

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
31 May 2001 (31.05.2001)

PCT

(10) International Publication Number
WO 01/38577 A211017 U.S. PTO
10/092885
03/06/02(51) International Patent Classification⁷: **C12Q 1/68**

Kenneth, W. [US/US]; 1403 Halkirk Way, BelAir, MD 21015 (US).

(21) International Application Number: PCT/US00/31922

(74) Agents: **HEMMENDINGER, Lisa, M. et al.**; Banner & Witcoff, Ltd., 11th floor, 1001 G Street, N.W., Washington, DC 20001-4597 (US).(22) International Filing Date:
21 November 2000 (21.11.2000)

(25) Filing Language: English

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(26) Publication Language: English

(30) Priority Data:
09/448,480 24 November 1999 (24.11.1999) US(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/448,480 (CON)
Filed on 24 November 1999 (24.11.1999)(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).(71) Applicant (*for all designated States except US*): **THE JOHNS HOPKINS UNIVERSITY** [US/US]; Suite 906, 111 Market Place, Baltimore, MD 21202 (US).**Published:**

— Without international search report and to be republished upon receipt of that report.

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **VELCULESCU, Victor, E.** [US/US]; Apt. 903, 1 E. Chase St., Baltimore, MD 21202 (US). **VOGELSTEIN, Bert** [US/US]; 3700 Breton Way, Baltimore, MD 21208 (US). **KINZLER,***For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: HUMAN TRANSCRIPTOMES

(57) Abstract: Global gene expression patterns have been characterized in normal and cancerous human cells using serial analysis of gene expression (SAGE). Cancer cell-specific, cell-type specific, and ubiquitously expressed genes have been identified. This information can be used to provide combination of cell type- and cancer-specific gene probes, as well as methods of using these probes to identify particular cell types, screen for useful drugs, reduce cancer-specific gene expression, standardize gene expression, and restore function to a diseased cell or tissue.

WO 01/38577 A2

WO 01/38577

PCT/US00/31922

HUMAN TRANSCRIPTOMES

This invention was made with government support under CA57345, CA62924, and CA43460 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

The characteristics of an organism are largely determined by the genes expressed within its cells and tissues. These expressed genes can be represented by transcriptomes that convey the identity and expression level of each expressed gene in a defined population of cells (1, 2). Although the entire sequence of the human genome will be elucidated in the near future (3), little is known about the many transcriptomes present in the human organism. Basic questions regarding the set of genes expressed in a given cell type, the distribution of expressed genes, and how these compare to genes expressed in other cell types, have remained largely unanswered.

General properties of gene expression patterns in eukaryotic cells were determined many years ago by RNA-cDNA reassociation kinetics (4), but these studies did not provide much information about the identities of the expressed genes within each expression class. Technological constraints have limited other analyses of gene expression to one or few genes at a time (5-9) or were non-quantitative (10, 11). Serial analysis of gene expression (SAGE) (12), one of several recently developed gene expression methods, has permitted the quantitative analysis of transcriptomes in the yeast *Saccharomyces cerevisiae* (1, 13). This effort identified the expression of known and previously unrecognized genes in *S.*

WO 01/38577

PCT/US00/31922

cerevisae (1, 14) and demonstrated that genome-wide expression analyses were practicable in eukaryotes.

Thus, there is a need in the art for the identification of transcriptomes which represent gene expression in particular cell types or under particular physiological conditions in eukaryotes, particularly in humans.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide such transcriptomes, individual polynucleotides, and methods of using the polynucleotides to identify particular cell types, screen for useful drugs, reduce cancer-specific gene expression, standardize gene expression, and restore function to a diseased cell or tissue. These and other objects of the invention are provided by one or more of the embodiments described below.

One embodiment of the invention is a method of identifying a cell as either a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, or a kidney epithelial cell. Expression in a test cell of a gene product of at least one gene is determined. The at least one gene comprises a sequence selected from at least one of the following groups:

- (a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;
- (b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;
- (c) the sequences shown in SEQ ID NOS:152-154 and 155;
- (d) the sequences shown in SEQ ID NOS:156-159 and 160;
- (e) the sequences shown in SEQ ID NOS:161-166 and 167;
- (f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;
- (g) the sequences shown in SEQ ID NOS:209 and 210; and
- (h) the sequences shown in SEQ ID NOS:211-224 and 225.

Expression of a gene product of at least one gene comprising a sequence shown in (a) identifies the test cell as a colon epithelial cell. Expression of a gene product of at least one gene comprising a sequence shown in (b) identifies the test cell as a brain cell. Expression

WO 01/38577

PCT/US00/31922

of a gene product of at least one gene comprising a sequence shown in (c) identifies the test cell as a keratinocyte. Expression of a gene product of at least one gene comprising a sequence shown in (d) identifies the test cell as a breast epithelial cell. Expression of a gene product of at least one gene comprising a sequence shown in (e) identifies the test cell as a lung epithelial cell. Expression of a gene product of at least one gene comprising a sequence shown in (f) identifies the test cell as a melanocyte. Expression of a gene product of at least one gene comprising a sequence shown in (g) identifies the test cell as a prostate cell. Expression of a gene product of at least one gene comprising a sequence shown in (h) identifies the test cell as a kidney epithelial cell.

Another embodiment of the invention is an isolated polynucleotide comprising a sequence selected from the group consisting of SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-84, 98, 103, 113, 115, 122, 129, 132, 134, 135, 140, 144, 149, 150, 153-168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-213, 216-224, 237, 239, 257, 263, 485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

Still another embodiment of the invention is a solid support comprising at least one polynucleotide. The polynucleotide comprises a sequence selected from at least one of the following groups:

- (a) the sequences shown in SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-83, and 84;
- (b) the sequences shown in SEQ ID NOS:98, 103, 113, 115, 122, 129, 132, 134, 135, 140, 144, 149, and 150;
- (c) the sequences shown in SEQ ID NOS:153-154 and 155;
- (d) the sequences shown in SEQ ID NOS:156-157 and 160;
- (e) the sequences shown in SEQ ID NOS:161-166 and 167;
- (f) the sequences shown in SEQ ID NOS:168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-207 and 208;
- (g) the sequences shown in SEQ ID NOS:209 and 210;

WO 01/38577

PCT/US00/31922

- (h) the sequences shown in SEQ ID NOS:211-213, 216-223, and 224;
- (i) the sequences shown in SEQ ID NOS:237, 239, 257, and 263; or
- (j) the sequences shown in SEQ ID NOS:485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

Even another embodiment of the invention is a method of identifying a test cell as a cancer cell. Expression in a test cell of a gene product of at least one gene is determined. The at least one gene comprises a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259-260, and 262-265. An increase in expression of at least two-fold relative to expression of the at least one gene in a normal cell identifies the test cell as a cancer cell.

Yet another embodiment of the invention is a method of reducing expression of a cancer-specific gene in a human cell. A reagent which specifically binds to an expression product of a cancer-specific gene is administered to the cell. The cancer-specific gene comprises a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259-260, and 262-265. Expression of the cancer-specific gene is thereby reduced relative to expression of the cancer-specific gene in the absence of the reagent.

Even another embodiment of the invention is a method for comparing expression of a gene in a test sample to expression of a gene in a standard sample. A first ratio and a second ratio are determined. The first ratio is an amount of an expression product of a test gene in a test sample to an amount of an expression product of at least one gene comprising a sequence selected from the group consisting of SEQ ID NOS:266-375, 377-652, 654-796, and 798-1448 in the test sample. The second ratio is an amount of an expression product of the test gene in a standard sample to an amount of an expression product of the at least one gene in the standard sample. The first and second ratios are compared. A difference between the first and second ratios indicates a difference in the amount of the expression product of the test gene in the test sample.

Still another embodiment of the invention is a method of screening candidate anti-cancer drugs. A cancer cell is contacted with a test compound. Expression of a gene

WO 01/38577

PCT/US00/31922

product of at least one gene in the cancer cell is measured. The at least one gene comprises a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259, 260, 262-263, and 265. A decrease in expression of the gene product in the presence of a test compound relative to expression of the gene product in the absence of the test compound identifies the test compound as a potential anti-cancer drug.

Still another embodiment of the invention is a method of screening test compounds for the ability to increase an organ or cell function. A selected from the group consisting of a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, and a kidney cell is contacted with a test compound. Expression in the cell of a gene product of at least one gene is measured. The gene comprises a sequence selected from at least one of the following groups:

- (a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;
- (b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;
- (c) the sequences shown in SEQ ID NOS:152-154 and 155;
- (d) the sequences shown in SEQ ID NOS:156-159 and 160;
- (e) the sequences shown in SEQ ID NOS:161-166 and 167;
- (f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207 and 208;
- (g) the sequences shown in SEQ ID NOS:209 and 210; and
- (h) the sequences shown in SEQ ID NOS:211-224 and 225.

An increase in expression of a gene product of at least one gene comprising a sequence shown in (a) identifies the test compound as a potential drug for increasing a function of a colon cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (b) identifies the test compound as a potential drug for increasing a function of a brain cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (c) identifies the test compound as a potential drug for increasing a function of a skin cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (d) identifies the test compound as a potential drug for increasing a function of a breast cell. An increase in expression of a gene product

WO 01/38577

PCT/US00/31922

of at least one gene comprising a sequence shown in (e) identifies the test compound as a potential drug for increasing a function of a lung cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (f) identifies the test compound as a potential drug for increasing a function of a melanocyte. An increase in expression of a gene product of at least one gene comprising a sequence shown in (g) identifies the test compound as a potential drug for increasing a function of a prostate cell. An increase in expression of a gene product of at least one gene comprising a sequence shown in (h) identifies the test compound as a potential drug for increasing a function of a kidney cell.

Yet another embodiment of the invention is a method to restore function to a diseased tissue. A gene is delivered to a diseased cell selected from the group consisting of a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, and a kidney cell. The gene comprises a nucleotide sequence selected from at least one of the following groups:

- (a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;
- (b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;
- (c) the sequences shown in SEQ ID NOS:152-154 and 155;
- (d) the sequences shown in SEQ ID NOS:156-159 and 160;
- (e) the sequences shown in SEQ ID NOS:161-166 and 167;
- (f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;
- (g) the sequences shown in SEQ ID NOS:209 and 210; and
- (h) the sequences shown in SEQ ID NOS:211-224 and 225.

Expression of the gene in the diseased cell is less than expression of the gene in a corresponding cell which is normal. If the diseased cell is a colon epithelial cell, then the nucleotide sequence is selected from (a). If the diseased cell is a brain cell, then the nucleotide sequence is selected from (b). If the diseased cell is a keratinocyte, then the nucleotide sequence is selected from (c). If the diseased cell is a breast epithelial cell, then the nucleotide sequence is selected from (d). If the diseased cell is a lung epithelial cell,

WO 01/38577

PCT/US00/31922

then the nucleotide sequence is selected from (e). If the diseased cell is a melanocyte, then the nucleotide sequence is selected from (f). If the diseased cell is a prostate cell, then the nucleotide sequence is selected from (g). If the diseased cell is a kidney cell, then the nucleotide sequence is selected from (h).

Thus, the invention provides transcriptomes, polynucleotides, and methods of identifying particular cell types, reducing cancer-specific gene expression, identifying cancer cells, standardizing gene expression, screening test compounds for the ability to increase an organ or a cell function, and restoring function to a diseased tissue.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. Sampling of gene expression in colon cancer cells. Analysis of transcripts at increasing increments of transcript tags indicates that the fraction of new transcripts identified approaches 0 at approximately 650,000 total tags.

FIG. 2. Colon cancer cell Rot curve.

FIGS. 3A-3C. Gene expression in different tissues. **FIG. 3A.** Fold reduction or induction of unique transcripts for each of the comparisons analyzed. The source of the transcripts included in each comparison are displayed in FIG. 3C. The relative expression of each transcript was determined by dividing the number of transcript tags in each comparison in the order displayed in FIG. 3C. To avoid division by 0, we used a tag value of 1 for any tag that was not detectable in one of the samples. We then rounded these ratios to the nearest integer; their distribution is plotted on the X axis. The number of transcripts displaying each ratio is plotted on the Y axis. Each comparison is represented by a specific color (see below or FIG. 3C). **FIG. 3B.** Expression of transcripts for each comparison, where values on X and Y axes represent the observed transcript tag abundances in each of the two compared sets. Light Blue symbols: DLD1 in different physiologic conditions; Yellow symbols: DLD1 cells (X axis) versus HCT116 cells (Y axis); Red symbols: colon cancer cells (X axis) versus normal brain (Y axis); and Dark Blue symbols: colon cancer cells (X axis) versus hemangiopericytoma (Y axis). **FIG. 3C.** Fraction of transcripts with dramatically altered expression. For each comparison, Expression Change denotes the number of transcripts induced or reduced 10 fold, and (%) denotes the number of altered

WO 01/38577

PCT/US00/31922

transcripts divided by the number of unique transcripts in each case. Differences between expression changes were evaluated using the chi squared test, where the expected expression changes were assumed to be the average expression change for any two comparisons.

TABLE LEGENDS

Table 1. Table of tissues and transcript tags analyzed. "Tissues" represents the source of the RNA analyzed, "Libraries" indicates the number of SAGE libraries analyzed, "Total Transcripts" is the total number of transcripts analyzed from each tissue, and "Unique Transcripts" denotes the number of unique transcripts observed in each tissue.

Table 2. Table of transcript abundance. "Copies/cell" denotes the category of expression level analyzed in transcript copies per cell, "Unique Transcripts" represents the number of unique transcripts observed and those matching GenBank genes or ESTs, and "Mass fraction mRNA" represents the fraction of mRNA molecules contained in each expression category.

Table 3. Table showing tissue-specific transcripts. The number in parentheses adjacent to the tissue type indicates the percent of transcripts exclusively expressed in a given tissue at 10 copies per cell. "Transcript tag" denotes the 10 bp tag adjacent to 4 bp *Nla*III anchoring enzyme site, "Copies/cell" denotes the transcript copies per cell expressed, and "UniGene Description" provides a functional description of each matching UniGene cluster (from UniGene Build No. 67). As UniGene cluster numbers change over time, the most recent cluster assignment for each tag can be obtained individually at <http://www.ncbi.nlm.nih.gov/SAGE/SAGetag.cgi> (Lal *et al.*, "A public database for gene expression in human cancers," *Cancer Research*, in press) or for the entire table at <http://www.sagenet.org/transcriptome>.

Table 4. Table showing ubiquitously expressed genes. "Copies/cell" denotes the average expression level of each transcript from all tissues examined, "Range" represents the range in expression for each transcript tag among all tissues analyzed in copies per cell, and "Range/Avg" is the ratio of the range to the average expression level and provides a measure of uniformity of expression. Other table columns are the same as in Table 5. The

WO 01/38577

PCT/US00/31922

entire table of uniformly expressed transcripts also is available at <http://www.sagenet.org/transcriptome>.

Table 5. Table showing transcripts uniformly elevated in human cancers. Transcripts expressed at 3 copies/cell whose expression is at least 2-fold higher in each cancer compared to its corresponding normal tissue. CC, colon cancer; BC, brain cancer; BrC, breast cancer; LC, lung cancer; M, melanoma; NC, normal colon epithelium; NB, normal brain; NBr, normal breast epithelium; NL, normal lung epithelium; NM, normal melanocytes. "Avg T/N" is the average ratio of expression in tumor tissue divided by normal tissue (for the purpose of obtaining this ratio, expression values of 0 are converted to 0.5). Other table columns are the same as in Table 5.

Table 6. Table showing transcripts expressed in colon cancer cells at a level of at least 500 copies per cell.

Table 7. Table showing transcripts expressed at a level of at least 500 copies per cell.

DETAILED DESCRIPTION OF THE INVENTION

It is a discovery of the present invention that particular sets of expressed genes ("transcriptomes") are expressed only in cancer cells; expression of these genes can be used, *inter alia*, to identify a test cell as cancerous and to screen for anti-cancer drugs. These cancer-specific genes can also provide targets for therapeutic intervention.

It is another discovery of the invention that other transcriptomes are differentially associated with distinct cell types; expression of genes of these transcriptomes can therefore be used to identify a test cell as belonging to one of these distinct cell types.

It is yet another discovery of the invention that genes of another transcriptome are expressed ubiquitously; expression of genes of this transcriptome can be used to standardize expression of other genes in a variety of gene expression assays.

To identify the transcriptomes described herein we used the SAGE method, as described in Velculescu *et al.* (1) and Velculescu *et al.* (12), to analyze gene expression in a variety of different human cell and tissue types. The SAGE method is also described in U.S. Patents 5,866,330 and 5,695,937. A total of 84 SAGE libraries were generated from

WO 01/38577

PCT/US00/31922

19 tissues (Table 1). Diseased tissues included cancers of the colon, pancreas, breast, lung, and brain, as well as melanoma, hemangiopericytoma, and polycystic kidney disease. Normal tissues included epithelia of the colon, breast, lung, and kidney, melanocytes, chondrocytes, monocytes, cardiomyocytes, keratinocytes, and cells of prostate and brain white matter and astrocytes.

A total of 3,496,829 transcript tags were analyzed and found to represent 134,135 unique transcripts after correcting for sequencing errors (transcript data available at <http://www.sagenet.org./transcriptome>). Expression levels for these transcripts ranged from 0.3 to a high of 9,417 transcript copies per cell in lung epithelium. Comparison against the GenBank and UniGene collections of characterized genes and expressed sequence tags (ESTs) revealed that 6,900 transcript tags matched known genes, while 65,735 matched ESTs. The remaining 61,500 transcript tags (46%) had no matches to existing databases and corresponded to previously uncharacterized or partially sequenced transcripts.

Each of the genes or transcripts whose expression can be measured in the methods of the invention comprises a unique sequence of at least 10 contiguous nucleotides (the "SAGE tag"). Genes which are differentially expressed in colon, lung, kidney, and breast epithelial cells, brain cells, prostate cells, keratinocytes, or melanocytes are shown in Table 3. Ubiquitously expressed genes are shown in Table 4. Transcripts which are expressed only in cancer tissues, *e.g.*, colon cancer, breast cancer, brain cancer, liver cancer, and melanoma, are shown in Table 5.

This information provides heretofore unavailable picture of human transcriptomes. These results, like the human genome sequence, provide basic information integral to future experimentation in normal and disease states. Because SAGE analyses provide absolute expression levels, future SAGE data can be directly integrated with those described here to provide progressively deeper insights into gene expression patterns. Eventually, a relatively complete description of the transcripts expressed in diverse cell types and in various physiologic states can be obtained.

Isolated polynucleotides

The invention provides isolated polynucleotides comprising either

WO 01/38577

PCT/US00/31922

deoxyribonucleotides or ribonucleotides. Isolated DNA polynucleotides according to the invention contain less than a whole chromosome and can be either genomic DNA or DNA which lacks introns, such as cDNA. Isolated DNA polynucleotides can comprise a gene or a coding sequence of a gene comprising a sequence as shown in SEQ ID NOS:1-1563, such as polynucleotides which comprise a sequence selected from the group consisting of SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-84, 98, 103, 113, 115, 122, 129, 132, 134, 135, 140, 144, 149, 150, 153-168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-213, 216-224, 237, 239, 257, 263, 485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

Any technique for obtaining a polynucleotide can be used to obtain isolated polynucleotides of the invention. Preferably the polynucleotides are isolated free of other cellular components such as membrane components, proteins, and lipids. They can be made by a cell and isolated, or synthesized using an amplification technique, such as PCR, or by using an automatic synthesizer. Methods for purifying and isolating polynucleotides are routine and are known in the art.

Isolated polynucleotides also include oligonucleotide probes, which comprise at least one of the sequences shown in SEQ ID NOS:1-1563. An oligonucleotide probe is preferably at least 10, 11, 12, 13, 14, 15, 20, 30, 40, or 50 or more nucleotides in length. If desired, a single oligonucleotide probe can comprise 2, 3, 4, or 5 or more of the sequences shown in SEQ ID NOS:1-1563. The probes may or may not be labeled. They may be used, for example, as primers for amplification reactions, such as PCR, in Southern or Northern blots, or for *in situ* hybridization.

Oligonucleotide probes of the invention can be made by expressing cDNA molecules comprising one or more of the sequences shown in SEQ ID NOS:1-1563 in an expression vector in an appropriate host cell. Alternatively, oligonucleotide probes can be synthesized chemically, for example using an automated oligonucleotide synthesizer, as is known in the art.

WO 01/38577

PCT/US00/31922

Solid Supports Comprising Polynucleotides

Polynucleotides, particularly oligonucleotide probes, preferably are immobilized on a solid support. A solid support can be any surface to which a polynucleotide can be attached. Suitable solid supports include, but are not limited to, glass or plastic slides, tissue culture plates, microtiter wells, tubes, gene "chips," or particles such as beads, including but not limited to latex, polystyrene, or glass beads. Any method known in the art can be used to attach a polynucleotide to a solid support, including use of covalent and non-covalent linkages, passive absorption, or pairs of binding moieties attached respectively to the polynucleotide and the solid support.

Polynucleotides are preferably present on an array so that multiple polynucleotides can be simultaneously tested for hybridization to polynucleotides present in a single biological sample. The polynucleotides can be spotted onto the array or synthesized *in situ* on the array. Such methods include older technologies, such as "dot blot" and "slot blot" hybridization (53, 54), as well as newer "microarray" technologies (55-58). A single array contains at least one polynucleotide, but can contain more than 100, 500, 1,000, 10,000, or 100,000 or more different probes in discrete locations.

Determining expression of a gene product

Each of the methods of the invention involves measuring expression of a gene product of at least one of the genes identified in Tables 3, 4, and 5 (SEQ ID NOS:1-1448). If desired, expression of gene products of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 50, 75, 100, 125, 250, 500, 1,000, 1,250, or more genes can be determined.

Either protein or RNA products of the disclosed genes can be determined. Either qualitative or quantitative methods can be used. The presence of protein products of the disclosed genes can be determined, for example, using a variety of techniques known to the art, including immunochemical methods such as radioimmunoassay, Western blotting, and immunohistochemistry. Alternatively, protein synthesis can be determined *in vivo*, in a cell culture, or in an *in vitro* translation system by detecting incorporation of labeled amino acids into protein products.

RNA expression can be determined, for example, using at least 1, 2, 3, 4, 5, 10, 15,

WO 01/38577

PCT/US00/31922

20, 25, 30, 50, 75, 100, 125, 250, 500, 1,000, 5,000, 10,000, or 100,000 or more oligonucleotide probes, either in solution or immobilized on a solid support, as described above. Expression of the disclosed genes is preferably determined using an array of oligonucleotide probes immobilized on a solid support. *In situ* hybridization can also be used to detect RNA expression.

Identification of Cell Types

Cell-type specific genes are expressed at a level greater than 10 copies per cell in a particular cell type, such as epithelial cells of the colon, breast, lung, and kidney, keratinocytes, melanocytes, and cells from the prostate and brain, but are not expressed in cells of other tissues. Such cell-type specific genes represent "cell-type specific transcriptomes." The fraction of cell-type-specific transcripts ranges from 0.05% in normal prostate to 1.76% in normal colon epithelium. Approximately 50% of these transcripts tags match known genes or ESTs. The vast majority of these cell-type-specific genes have not been previously reported in the literature to be cell-type specific.

Cell type-specific genes are shown in Table 3. Genes which comprise the sequences shown in SEQ ID NOS:1-85 are uniquely expressed in colon epithelial cells. Genes which comprise the sequences shown in SEQ ID NOS:86-151 are uniquely expressed in brain cells. Genes which comprise the sequences shown in SEQ ID NOS:152-155 are uniquely expressed in keratinocytes. Genes which comprise the sequences shown in SEQ ID NOS:156-160 are uniquely expressed in breast epithelial cells. Genes which comprises the sequences shown in SEQ ID NOS:161-167 are uniquely expressed in lung epithelial cells. Genes which comprises the sequences shown in SEQ ID NOS:168-208 are uniquely expressed in melanocytes. Genes which comprise the sequences shown in SEQ ID NOS:209 and 210 are uniquely expressed in prostate cells. Genes which comprise the sequences shown in SEQ ID NOS:211-225 are uniquely expressed in kidney epithelial cells. Thus, determination of expression of at least one gene from each of these uniquely expressed groups, particularly those not previously known to be uniquely expressed, can be used to identify a test cell as an epithelial cell of the colon, breast, lung, and kidney, a keratinocyte, a melanocyte, or a cell from the prostate or brain.

WO 01/38577

PCT/US00/31922

Test cells can be obtained, for example, from biopsy or surgical samples, forensic samples, cell lines, or primary cell cultures. Test cells include normal as well as cancer cells, such as primary or metastatic cancer cells.

To identify a test cell as an epithelial cell of the colon, breast, lung, and kidney, a keratinocyte, a melanocyte, or a cell from the prostate or brain, expression of a gene product of at least one gene is determined, using methods such as those described above. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:2, 5-18, and 20-85, the test cell is identified as a colon epithelial cell. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, and 131-151, the test cell is identified as a brain cell. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:152-155, the test cell is identified as a keratinocyte. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:156-160, the test cell is identified as a breast epithelial cell. If a test cell expresses a gene comprising a sequence shown in SEQ ID NOS:161-167, the test cell is identified as a lung epithelial cell. Expression of a gene comprising a sequence shown in SEQ ID NOS:168, 170, 172-177, 179-188, and 190-208 identifies the test cell as a melanocyte. Expression of a gene comprising a sequence shown in SEQ ID NOS:209 and 210 identifies the test cell as a prostate cell. Expression of a gene which comprises a sequence shown in SEQ ID NOS:211-225 identifies the test cell as a kidney epithelial cell.

Identifying a Test Cell as a Cancer Cell

A cancer-specific gene is expressed at a level of at least 3 copies per cancer cell, such as a colon cancer, breast cancer, brain cancer, lung cancer, or melanoma cell, at a level which is at least two-fold higher than expression of the same gene in a corresponding normal cell. Cancer-specific genes which comprise the sequences shown in SEQ ID NOS:226-265 (Table 5) represent a "cancer transcriptome." SEQ ID NOS:237, 239, 257, and 263 are sequences which are found in transcripts of novel cancer-specific genes of the invention. Oligonucleotide probes corresponding to cancer-specific genes can be used, for example, to detect and/or measure expression of cancer-specific genes for diagnostic purposes, to assess efficacy of various treatment regimens, and to screen for potential anti-

WO 01/38577

PCT/US00/31922

cancer drugs.

For example, determination of the expression level of any of these genes in a test cell relative to the expression level of the same gene in a normal cell (a cell which is known not to be a cancer cell) can be used to determine whether the test cell is a cancer cell or a non-cancer cell.

Test cells can be any human cell suspected of being a cancer cell, including but not limited to a colon epithelial cell, a breast epithelial cell, a lung epithelial cell, a kidney epithelial cell, a melanocyte, a prostate cell, and a brain cell. Test cells can be obtained, for example, from biopsy samples, surgically excised tissues, forensic samples, cell lines, or primary cell cultures. Comparison can be made to a non-cancer cell type, including to the corresponding non-cancer cell type, either at the time expression is measured in the test cell or by reference to a previously determined expression standard.

To identify a test cell as a cancer cell, expression of a gene product of at least one gene is determined, using methods such as those described above. The at least one gene comprises a sequence selected from the group consisting of SEQ ID NOS:226-265, particularly from the group consisting of SEQ ID NOS:228, 230-236, 238, 240-256, 258-260, and 262-265. An increase in expression of the at least one gene in the test cell which is at least two-fold more than the expression of the at least one gene in a cell which is not cancerous identifies the test cell as a cancer cell.

Reducing Cancer-Specific Gene Expression

Cancer-specific genes provide potential therapeutic targets for treating cancer or for use in model systems, for example, to screen for agents which will enhance the effect of a particular compound on a potential therapeutic target. Thus, a reagent can be administered to a human cell, either *in vitro* or *in vivo*, to reduce expression of a cancer-specific gene. The reagent specifically binds to an expression product of a gene comprising a sequence selected from the group consisting of SEQ ID NOS:226-265, particularly from the group consisting of SEQ ID NOS:228, 230-236, 238, 240-256, 258-260, and 262-265.

If the expression product is a protein, the reagent is preferably an antibody. Protein products of cancer-specific genes can be used as immunogens to generate antibodies, such

WO 01/38577

PCT/US00/31922

as a polyclonal, monoclonal, or single-chain antibodies, as is known in the art. Protein products of cancer-specific genes can be isolated from primary or metastatic tumors, such as primary colon adenocarcinomas, lung cancers, astrocytomas, glioblastomas, breast cancers, and melanomas. Alternatively, protein products can be prepared from cancer cell lines such as SW480, HCT116, DLD1, HT29, RKO, 21-PT, MDA-468, A549, and the like. If desired, cancer-specific gene coding sequences can be expressed in a host cell or in an *in vitro* translation system. An antibody which specifically binds to a protein product of a cancer-specific gene provides a detection signal at least 5-, 10-, or 2-fold higher than a detection signal provided with other proteins when used in an immunochemical assay. Preferably, the antibody does not detect other proteins in immunochemical assays and can immunoprecipitate the cancer-specific protein product from solution.

For administration *in vitro*, an antibody can be added to a tissue culture preparation, either as a component of the medium or in addition to the medium. In another embodiment, antibodies are delivered to specific tissues *in vivo* using receptor-mediated targeted delivery. Receptor-mediated DNA delivery techniques are taught in, for example, Findeis *et al. Trends in Biotechnol.* 11, 202-05, (1993); Chiou *et al.*, GENE THERAPEUTICS: METHODS AND APPLICATIONS OF DIRECT GENE TRANSFER (J.A. Wolff, ed.) (1994); Wu & Wu, *J. Biol. Chem.* 263, 621-24, 1988; Wu *et al.*, *J. Biol. Chem.* 269, 542-46, 1994; Zenke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 87, 3655-59, 1990; Wu *et al.*, *J. Biol. Chem.* 266, 338-42, 1991.

If single-chain antibodies are used, polynucleotides encoding the antibodies can be constructed and introduced into cells using well-established techniques including, but not limited to, transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, "gene gun," and DEAE- or calcium phosphate-mediated transfection.

Effective *in vivo* dosages of an antibody are in the range of about 5 μ g to about 50 μ g/kg of patient body weight, about 50 μ g to about 5 mg/kg, about 100 μ g to about 500 μ g/kg of patient body weight, and about 200 to about 250 μ g/kg. For administration of polynucleotides encoding single-chain antibodies, effective *in vivo* dosages are in the range

WO 01/38577

PCT/US00/31922

of about 100 ng to about 200 ng, 500 ng to about 50 mg, about 1 μ g to about 2 mg, about 5 μ g to about 500 μ g, and about 20 μ g to about 100 μ g of DNA.

If the expression product is mRNA, the reagent is preferably an antisense oligonucleotide. The nucleotide sequence of an antisense oligonucleotide is complementary to at least a portion of the sequence of the cancer-specific gene. Preferably, the antisense oligonucleotide sequence is at least 10 nucleotides in length, but can be at least 11, 12, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides long. Longer sequences also can be used. An antisense oligonucleotide which specifically binds to an mRNA product of a cancer-specific gene preferably hybridizes with no more than 3 or 2 mismatches, preferably with no more than 1 mismatch, even more preferably with no mismatches.

Antisense oligonucleotides can be deoxyribonucleotides, ribonucleotides, or a combination of both. Oligonucleotides, including modified oligonucleotides, can be prepared by methods well known in the art (47-52) and introduced into human cells using techniques such as those described above. The cells can be in a primary culture of human tumor cells, in a human tumor cell line, or can be primary or metastatic tumor cells present in a human body.

Preferably, a reagent reduces expression of a cancer-specific gene by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80% relative to expression of the gene in the absence of the reagent. Most preferably, the level of gene expression is decreased by at least 90%, 95%, 99%, or 100%. The effectiveness of the mechanism chosen to decrease the level of expression of a cancer-specific gene can be assessed using methods well known in the art, such as hybridization of nucleotide probes to cancer-specific gene mRNA, quantitative RT-PCR, or immunologic detection of a protein product of the cancer-specific gene.

Screening for Anti-Cancer Drugs

According to the invention, test compounds can be screened for potential use as anti-cancer drugs by assessing their ability to suppress or decrease the expression of at least one cancer-specific gene. The cancer-specific gene comprises a sequence selected from the group consisting of SEQ ID NOS:226-265, particularly from the group consisting of SEQ

WO 01/38577

PCT/US00/31922

ID NOS:228, 230-236, 238, 240-256, 258-260, and 262-265. Test compounds can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity, including small molecules from compound libraries. Test substances can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants, or can be produced recombinantly or synthesized by chemical methods known in the art.

To screen a test compound for use as a possible anti-cancer drug, a cancer cell is contacted with the test compound. The cancer cell can be a cell of a primary or metastatic tumor, such as a tumor of the colon, breast, lung, prostate, brain, or kidney, or a melanoma, which is isolated from a patient. Alternatively, a cancer cell line, such as colon cancer cell lines HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO, breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474, the A549 lung cancer cell line, and the H392 glioblastoma cell line, can be used.

Expression of a gene product of at least one gene is determined using methods such as those described above. The gene comprises a sequence selected from the group consisting of SEQ ID NOS:226-265, preferably from the group consisting of SEQ ID NOS:228, 230-236, 238, 240-256, 258-260, and 262-265, even more preferably from the group consisting of SEQ ID NOS:237, 239, 257, and 263. A decrease in expression of the gene in the cancer cell identifies the test compound as a potential anti-cancer drug.

Standardizing Expression of a Test Gene

Genes which comprise the sequences shown in SEQ ID NOS:266-1448 (Table 4) are expressed at a level of at least five transcript copies per cell in every cell type analyzed, including epithelia of the colon, breast, lung, and kidney, melanocytes, chondrocytes, monocytes, cardiomyocytes, keratinocytes, prostate cells, and astrocytes, oligodendrocytes, and other cells present in the white matter of brain. These genes thus represent members of the "minimal transcriptome," the set of genes expressed in all human cells. The minimal transcriptome includes well known genes which are often used as experimental controls to normalize gene expression, such as glyceraldehyde 3-phosphate dehydrogenase, elongation factor 1 alpha, and gamma actin.

WO 01/38577

PCT/US00/31922

Ubiquitously expressed genes can be used to compare expression of a test gene in a test sample to expression of a gene in a standard sample. A ubiquitously expressed gene preferably comprises a sequence shown in SEQ ID NOS:266-375, 377-652, 654-796, and 798-1448, and more preferably comprises a sequence shown in SEQ ID NOS:282, 288, 300, 302, 308, 320, 323, 363, 368, 379, 381, 444, 453, 518, 531, 535, 538, 542, 579, 580, 594, 600, 604, 617, 626, 641, 650, 717, 728, 776, 777, 794, 818, 822, 842, 885, 887, 899, 900, 902, 904, 914, 930, 960, 964, 1001, 1015, 1020, 1027, 1035, 1090, 1113, 1119, 1146, 1151, 1163, 1233, 1235, 1252, 1255, 1270, 1340, 1345, 1356, 1359, 1360, 1362, 1385, 1415, and 1441.

Two ratios are determined using gene expression assays such as those described above. The first ratio is an amount of an expression product of a test gene in a test sample to an amount of an expression product of at least one ubiquitously expressed gene comprising a sequence selected from the group consisting of SEQ ID NOS:266-375, 377-652, 798-1447, and 1448 in the test sample. The second ratio is an amount of an expression product of the test gene in a standard sample to an amount of an expression product of the ubiquitously expressed gene in the standard sample. Expression of either the test gene or the ubiquitously expressed gene can be used as the denominator. If desired, multiple ratios can be determined, such as (a) an amount of an expression product of more than one test gene to that of a single ubiquitously expressed gene, (b) an amount of an expression product of a single test gene to that of more than one ubiquitously expressed genes, or (c) an amount of an expression product of more than one test gene to that of more than one ubiquitously expressed gene. Optionally, the ratio in the standard sample can be pre-determined.

The ratios determined in the test and standard samples are compared. A difference between the ratios indicates a difference in the amount of the expression product of the test gene in the test sample.

The standard and test samples can be matched samples, such as whole cell cultures or homogenates of cells (such as a biopsy sample) and differ only in that the test biological sample has been subjected to a different environmental condition, such as a test compound, a drug whose effect is known or unknown, or altered temperature or other environmental

WO 01/38577

PCT/US00/31922

condition. Alternatively, the test and standard samples can be corresponding cell types which differ according to developmental age. In one embodiment, the test sample is a cancer cell, such as a colon cancer, breast cancer, lung cancer, melanoma, or brain cancer cell, and the standard sample is a normal cell.

The test gene can be a gene which encodes a protein whose biological function is known or unknown. Preferably the ratio of expression between the test gene and expression of the ubiquitously expressed gene is consistent in the standard sample. Even more preferably, expression of the ubiquitously expressed gene is not altered in the test sample. A difference between the first ratio of expression in the test sample and a second ratio of expression in the standard sample can therefore be used to indicate a difference in expression of the test gene in the test sample.

Screening for Compounds for Increasing an Organ or Cell Function

Test compounds can be screened for the ability to increase an organ or cell function by assessing their ability to increase expression of at least one tissue-specific gene. The tissue-specific gene comprises a sequence selected from at least one of the following groups:

- (a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;
- (b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;
- (c) the sequences shown in SEQ ID NOS:152-154, and 155;
- (d) the sequences shown in SEQ ID NOS:156-159 and 160;
- (e) the sequences shown in SEQ ID NOS:161-166 and 167;
- (f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;
- (g) the sequences shown in SEQ ID NOS:209 and 210; and
- (h) the sequences shown in SEQ ID NOS:211-224 and 225.

As with the anti-cancer drug screening method described above, test compounds can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity, including small molecules from compound libraries.

WO 01/38577

PCT/US00/31922

Test substances can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants, or can be produced recombinantly or synthesized by chemical methods known in the art.

To screen a test compound for the ability to increase an organ or cell function, a cell, such as a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, or a kidney cell, is contacted with the test compound. The cell can be a primary culture, such as an explant culture, of tissue obtained from a human, or can originate from an established cell line.

Expression of a gene product of at least one gene is determined using methods such as those described above. An increase in expression of a gene product of at least one gene comprising a sequence selected from (a) identifies the test compound as a potential drug for increasing a function of a colon cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (b) identifies the test compound as a potential drug for increasing a function of a brain cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (c) identifies the test compound as a potential drug for increasing a function of a skin cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (d) identifies the test compound as a potential drug for increasing a function of a breast cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (e) identifies the test compound as a potential drug for increasing a function of a lung cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (f) identifies the test compound as a potential drug for increasing a function of a melanocyte. An increase in expression of a gene product of at least one gene comprising a sequence selected from (g) identifies the test compound as a potential drug for increasing a function of a prostate cell. An increase in expression of a gene product of at least one gene comprising a sequence selected from (h) identifies the test compound as a potential drug for increasing a function of a kidney cell.

Restoring Function to a Diseased Tissue or Cell

Function can be restored to a diseased tissue or cell, such as a melanocyte or a colon,

WO 01/38577

PCT/US00/31922

brain, keratinocyte, breast, lung, prostate, or kidney cell, by delivering an appropriate tissue-specific gene to cells of that tissue. The tissue specific gene comprises a nucleotide sequence selected from at least one of the following groups:

- (a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85 (colon-specific);
- (b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151 (brain-specific);
- (c) the sequences shown in SEQ ID NOS:152-154, and 155 (keratinocyte-specific);
- (d) the sequences shown in SEQ ID NOS:156-159 and 160 (breast-specific);
- (e) the sequences shown in SEQ ID NOS:161-166 and 167 (lung-specific);
- (f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208 (melanocyte-specific);
- (g) the sequences shown in SEQ ID NOS:209 and 210 (prostate-specific); and
- (h) the sequences shown in SEQ ID NOS:211-224 and 225 (kidney-specific).

Expression of the gene in a cell of the diseased tissue preferably is 10, 20, 30, 40, 50, 60, 70, 80, or 90% less than expression of the gene in a cell of the corresponding tissue which is normal. In some cases, the diseased cell fails to express the gene. A tissue-specific gene which is administered to cells for this purpose includes a polynucleotide comprising a coding sequence which is intron-free, such as a cDNA, as well as a polynucleotide which comprises elements in addition to the coding sequence, such as regulatory elements.

Coding sequences of many of the tissue-specific genes disclosed herein are publicly available. For the novel tissue-specific genes identified here, coding sequences can be obtained using a variety of methods, such as restriction-site PCR (Sarkar, *PCR Methods Applic.* 2:318-322, 1993), inverse PCR (Triglia *et al.*, *Nucleic Acids Res.* 16:8186, 1988), capture PCR (Lagerstrom, *et al.*, *PCR Methods Applic.* 1:111-119, 1991). Alternatively, the partial sequences disclosed herein can be nick-translated or end-labeled with ³²P using polynucleotide kinase using labeling methods known to those with skill in the art (BASIC METHODS IN MOLECULAR BIOLOGY, Davis *et al.*, eds., Elsevier Press, N.Y., 1986). A lambda library prepared from the appropriate human tissue can then be directly screened with the labelled sequences of interest.

Many methods for introducing polynucleotides into cells or tissues are available and

WO 01/38577

PCT/US00/31922

can be used to deliver a tissue-specific gene to a cell *in vitro* or *in vivo*. Introduction of the tissue-specific gene into a cell can be accomplished by any method by which a nucleic acid molecule can be inserted into a cell, such as transfection, electroporation, microinjection, lipofection, adsorption, and protoplast fusion. For *in vitro* administration, a tissue-specific gene can be added to a tissue culture preparation, either as a component of the medium or in addition to the medium. *In vivo* administration can be by means of direct injection of a vector comprising a tissue-specific gene to the particular tissue or cells to which the tissue-specific gene is to be delivered. Alternatively, the tissue-specific gene can be included in a vector which is capable of targeting a particular tissue and administered systemically (59-61).

For *in vitro* administration, suitable concentrations of a tissue-specific gene in the culture medium range from at least about 10 pg to 100 pg/ml, about 100 pg to about 500 pg/ml, about 500 pg to about 1 ng/ml, about 1 ng to about 10 ng/ml, about 10 ng to about 100 ng/ml, or about 100 ng/ml to about 500 ng/ml. For local administration, effective dosages of a tissue-specific gene range from at least about 10 ng to about 100 ng, about 50 ng to 150 ng, about 100 ng to about 250 ng, about 1 µg to about 10 µg, about 5 µg to about 50 µg, about 25 µg to about 100 µg, about 75 µg to about 250 µg, about 100 µg to about 250 µg, about 200 µg to about 500 µg, about 500 µg to about 1 mg, about 1 mg to about 10 mg, about 5 mg to about 50 mg, about 25 mg to about 100 mg, or about 50 mg to about 200 mg of DNA per injection. Suitable concentrations for systemic administration range from at least about 500 ng to about 50 mg, about 1 µg to about 2 mg, about 5 µg to about 500 µg, and about 20 µg to about 100 µg of DNA per kg of body weight.

Recombinant DNA technologies can be used to improve expression of the tissue-specific gene by manipulating, for example, the number of copies of the gene in the cell, the efficiency with which the gene is transcribed, the efficiency with which the resultant transcripts are translated, and the efficiency of post-translational modifications. Recombinant techniques useful for increasing the expression of a tissue-specific gene in a cell include, but are not limited to, providing the tissue-specific gene in a high-copy number plasmid, integrating the tissue-specific gene into one or more host cell chromosomes, adding vector stability sequences to plasmids, substituting or modifying

WO 01/38577

PCT/US00/31922

transcription control signals (e.g., promoters, operators, enhancers), substituting or modulating translational control signals (e.g., ribosome binding sites, Shine-Dalgarno sequences), and deleting sequences that destabilize transcripts. (See Dow *et al.*, U.S. Patent 5,935,568).

Preferably, delivery of the tissue-specific gene increases expression of a gene product of the tissue-specific gene in the cell or tissue by at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 98, 99, or 100% relative to expression of the tissue-specific gene in a diseased cell or tissue to which the gene has not been delivered. Expression of a protein product of the tissue-specific gene can be determined immunologically, using methods such as radioimmunoassay, Western blotting, and immunohistochemistry. Alternatively, incorporation of labeled amino acids into a protein product can be determined. RNA expression is preferably determined using one or more oligonucleotide probes, either in solution or immobilized on a solid support, as described above.

All documents cited in this disclosure are expressly incorporated herein. The above disclosure generally describes the present invention, and all references cited in this disclosure are incorporated by reference herein. A more complete understanding can be obtained by reference to the following specific examples which are provided for purposes of illustration only and are not intended to limit the scope of the invention.

EXAMPLE 1

Tissue samples and the SAGE method

RNA for normal tissues was obtained from the following sources: colon epithelial cells isolated from sections of normal colon mucosa from two patients (41); HaCaT keratinocyte cells (42), normal mammary epithelial cells from two individuals (Clonetics); normal bronchial epithelial cell from two individuals (43); normal melanocytes from two individuals (Cascade Biologics); normal cultured monocytes, dendritic cells and TNF activated dendritic cells; two normal kidney epithelial cell lines; cultured chondrocyte cells from two normal individuals and one patient with osteoarthritic disease; normal fetal cardiomyocytes in normoxic and hypoxic conditions; and normal brain white matter from

WO 01/38577

PCT/US00/31922

two patients and normal cultured astrocyte cells.

RNA for diseased tissues was obtained from the following sources: primary colon adenocarcinomas from two patients, HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO colon cancer cell lines cultured *in vitro* in a variety of different cellular conditions including log phase growth, G1/G2 phase growth arrest, and apoptosis (40, 41, 44, 45); primary pancreatic adenocarcinomas from two patients and ASPC-1 and PL-45 pancreatic cancer cell lines (41); breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474; primary lung squamous cell cancers from two patients (43), primary lung adenocarcinoma from one patient, and the A549 lung cancer cell line (43); primary melanomas from 3 patients; kidney epithelial cells lines from two patients with polycystic kidney disease; hemangiopericytomas from 5 patients; primary glioblastoma tumors from two patients; and the H392 glioblastoma cell line.

Isolation of polyadenylate RNA and the SAGE method for all tissues was performed as previously described (1, 12; *see also* U.S. Patents 5,866,330 and 5,695,937).

EXAMPLE 2

Data analysis

The SAGE software (12) was used to analyze raw sequence data and to identify a total of 3,668,175 SAGE tags. Of these, 171,346 tags (4.7%) corresponded to linker sequences and were removed from further analysis. The remaining 3,496,829 tags were derived from transcript sequences, but a small fraction of these contained sequencing errors. SAGE analysis of yeast (1), for which the entire genome sequence is known, demonstrated a sequencing error rate of ~0.7% per bp, translating to a tag error rate of 6.8% (1-0.993; 10), in accord with sequence errors measured in the current data set.

To provide as accurate an estimate of unique genes as possible, we accounted for sequencing errors in two ways. First, we only considered tags that occurred twice in the data set. Although this requirement might have removed legitimate transcript tags expressed at very low levels (less than approximately 0.2 copies per cell, or 2 copies in 3,496,829 transcript tags), it eliminated the majority of sequencing errors (172,276 tags).

Second, because of the size of the data set utilized, it was possible that the same

WO 01/38577

PCT/US00/31922

sequencing error in a given tag may be observed multiple times. To account for these, tags with expression levels high enough to give multiple redundant errors were analyzed for single base substitutions, insertions, and deletions. If the observed expression level of a tag did not exceed its expected incidence due to redundant errors by a factor of five, it was assumed to be the result of a repeated sequencing error. This identified and removed an additional 27,051 unique tags (156,174 total tags), a number very similar to estimates of multiple sequencing errors obtained by Monte Carlo simulations.

In total, these corrections amount to a sequencing error rate of approximately 9.4%, suggesting that our analyses more than fully accounted for sequencing errors and that the remaining 134,135 unique transcript tags represented a conservative accounting of legitimate transcripts.

Transcript tags were matched to known genes and ESTs by use of tables containing matching 10 bp transcript sequences, UniGene clusters, GenBank accession numbers, and functional descriptions downloaded from the SAGEmap web site (<http://www.ncbi.nlm.nih.gov/SAGE>) (Lal *et al.*, in press) on Feb 23, 1999 (UniGene build 70, <http://www.ncbi.nlm.nih.gov/UniGene>), and the Microsoft Access software. As UniGene clusters numbers may change over time, the most recent tag to cluster mapping can be obtained for each transcript tag individually at <http://www.ncbi.nlm.nih.gov/SAGE/SAGETag.cgi>, or for the entire data set at <http://www.sagenet.org/transcriptome>. A total of 37,534 distinct transcripts from the UniGene database contained polyadenylation signals or polyadenylated tails and matched the collection of SAGE transcript tags; these corresponded to 23,534 unique UniGene clusters.

Transcript abundance per cell was determined simply by dividing the observed number of tags for a given transcript by the total number of transcripts obtained. An estimate of about 300,000 transcripts per cell was used to convert the abundances to copies per cell (46). For tissue specific transcripts, only transcript tags expressed at nominally ≥ 10 transcript copies per cell were considered in order to normalize for tissues with fewer total tags analyzed.

The following transcript data from this analysis are available electronically at the

WO 01/38577

PCT/US00/31922

SAGEnet web site (<http://www.sagenet.org/transcriptome>) with the corresponding expression levels and UniGene descriptions: 134,135 unique transcript tags identified from 3.5 million total transcripts tags; 69,381 transcript tags identified from colon cancer cells; 217 transcripts that are exclusively expressed in colon epithelium, keratinocytes, breast epithelium, lung epithelium, melanocytes, kidney epithelium and cells from prostate and brain; 987 transcripts that were expressed in all tissues. Individual transcript libraries from a total of ~800,000 transcript tags from colon epithelium, normal brain, colon cancer, and brain cancer are available at the SAGEmap web site (<http://www.ncbi.nlm.nih.gov/SAGE>) (Lal *et al.*, in press).

EXAMPLE 3

Estimation of the number of genes present in the human genome

The transcripts detected by SAGE provides an estimate of the number of genes present in the human genome. Historically, estimates of the number of unique genes in the genome have ranged from 60,000 to over 100,000 genes using analyses of EST clustering (15), frequency of genes in characterized genomic regions, frequency of CpG islands (16), and RNA-cDNA reassociation kinetics (4). If one were to assume that each unique transcript tag observed by SAGE corresponded to a unique gene, our data would indicate that there are approximately 134,000 genes in the human genome.

However, such an approach is likely to overestimate the number of unique genes in the genome, as distinct transcripts can be derived from a single gene. Multiple sites for polyadenylation (17), alternative splicing, premature transcriptional termination (18), as well as polymorphisms in the SAGE tag or nearby restriction endonuclease site could lead to multiple transcript tags for any one gene. An analysis of all publicly available 3' end-derived ESTs revealed that this was the case for many transcripts, and provided an estimate of the multiplicity of transcripts expected for individual genes. 37,534 distinct 3' transcripts containing polyadenylation signals or polyadenylated tails were observed to correspond to 23,534 unique UniGene clusters, an average 1.6 different transcripts per gene. Applying a similar calculation to our SAGE data would suggest that the 134,135 transcripts observed corresponded to 84,103 unique genes. As our SAGE data is by no

WO 01/38577

PCT/US00/31922

means a complete analysis of transcripts from all possible tissues, this estimate would provide a lower boundary for the number of unique genes in the genome. This figure is significantly higher than the 65,538 genes estimated from a clustering of 982,808 ESTs (UniGene Build 70) (15), and suggests that a substantial number of genes expressed at low levels may not be present in current EST databases.

EXAMPLE 4

Assessment of transcriptome complexity

Assessment of transcriptome complexity requires a relatively complete sampling of a transcriptome for the cell type under analysis. Human cells are thought to contain close to 300,000 mRNA molecules, and therefore an analysis of at least several hundred thousand transcripts would be needed. Approximately 350,000 and 300,000 transcripts were analyzed from DLD1 and HCT116 colorectal cancer cells, respectively. As these cancer cells are diploid, have similar genetic and phenotypic properties, and have very similar gene expression patterns (see below), transcript tags obtained from these cells were analyzed in combination as well as individually.

Analysis of either cell line afforded approximately a one fold coverage of the 300,000 mRNA molecules in a cell, while the combined set represented a two fold coverage even for mRNA molecules present at a single copy per cell. Measurement of ascertained new tags at increasing increments of tags indicated that the fraction of new transcripts from analysis of additional tags approached 0 at approximately 650,000 tags in the combined set (FIG. 1). This suggested that generation of further SAGE tags would yield few additional genes, and Monte Carlo simulations indicated that analysis of 643,283 tags would identify at least one tag for a given transcript 96% of the time if its expression level was at least two transcript copies per cell, and 83% of the time if its expression level was at least one transcript copy per cell.

The combined 643,283 transcript tags represented 69,381 unique transcripts, of which 44,174 corresponded to known genes or ESTs in the GenBank or UniGene databases while 25,207 represented previously undescribed transcripts (Table 2). Even when accounting for multiple unique transcripts per gene, these transcripts would represent at least 43,502

WO 01/38577

PCT/US00/31922

unique genes. This is substantially higher than the previous estimate of 15,000-25,000 expressed genes obtained by RNA-DNA reassociation kinetics in a variety of human cell types (4), and suggests that a significant fraction of the genome may be expressed in individual cell types. As the kinetics of reassociation of a particular class of RNA and cDNA may be affected by a number of experimental variables and may underestimate transcripts of low abundance (4), it is not surprising that our studies have detected a higher number of expressed genes than estimated by hybridization analysis in both human cells (Table 2) and yeast.

EXAMPLE 5

Expression levels of transcripts in colon cancer cells

Expression levels of transcripts in the colon cancer cell ranged from 0.5 to 2341 copies per cell. The 61 transcripts expressed at over 500 transcript copies per cell made up nearly 1/4 of the mRNA mass of the cell and the most highly expressed 623 genes accounted for 1/2 of the mRNA content. In contrast, the vast majority of unique transcripts were expressed at low levels, with just under 23% of the mRNA mass of the cell comprising 90% of the unique transcripts expressed (Table 2). A "virtual rot" analysis of the expressed transcripts identified a relatively continuous distribution of gene expression without markedly discrete abundance classes, similar to those observed in previous rot studies of human cancer cells (20) (FIG. 2).

The identities of the expressed genes reveal the diversity of expression of a human transcriptome (data available at <http://www.sagenet.org/transcriptome>). For example, highly expressed genes often encoded proteins important in protein synthesis, energy metabolism, cellular structure and certain tissue specific functions. Moderate and low abundance genes accounted for a multitude of cellular processes including protein modification enzymes, DNA replication machinery, cell surface receptors, components of signal transduction pathways and transcription factors as well as many other transcripts with currently unknown functions.

WO 01/38577

PCT/US00/31922

EXAMPLE 6*Differences in gene expression between different tissues*

Differences in gene expression between different tissues may provide insights into the specialized processes underlying human physiology in normal and diseased states. In line with previous observations, overall gene expression patterns among the 19 different tissues analyzed were similar (examples in FIGS. 3A-3C). Changes in gene expression between physiologic states of a particular cell type or between patient samples of the same tissue were less than changes between cell types of different origins (FIGS. 3A-3C). Likewise, only a small fraction of transcripts was exclusively expressed in a particular normal or disease tissue. Detailed analyses of transcripts from epithelia of colon, breast, lung, and kidney, melanocytes, and cells from prostate and brain, identified transcripts that were nominally expressed at greater than 10 copies per cell in one tissue but not in any other tissue studied. The fraction of these tissue-specific transcripts ranged from 0.05% in normal prostate to 1.76% in normal colon epithelium (Table 3). Approximately 50% of these transcript tags matched known genes or ESTs (examples in Table 3 and data available at <http://www.sagenet.org/transcriptome>). Some of these transcripts identified genes already reported to be important for tissue specific processes. For example, brain specific transcripts such as GABA receptor, myelin basic protein, and synaptopodin are known to be important for synaptic transmission (21) formation and maintenance of the myelin sheath (22) and dendrite shape and motility (23), respectively. Likewise, guanylin/uroguanylin (24), carbonic anhydrase 1 (25), and CDX2 (26) are known to be expressed in colonic epithelium. 5,6-dihydroxyindole-2-carboxylic acid oxidase has been shown to have an important role for normal melanocyte pigment synthesis (27), while expression of MART-1 and melastatin may have clinical implications for melanoma patients (28, 29). However, the vast majority of the tissue specific transcripts observed have not been previously reported in the literature and their roles in the tissues examined remain to be elucidated.

WO 01/38577

PCT/US00/31922

EXAMPLE 7*Minimal transcriptome*

Nearly 1000 transcripts were detected that were expressed at 5 transcript copies per cell in every cell type analyzed. These expressed genes represent a view into the “minimal transcriptome,” the set of genes expressed in all human cells. Such genes, listed in order of their uniformity of expression in Table 4 (and available at <http://www.sagenet.org/transcriptome>), largely represent well known constitutive or housekeeping genes thought to provide the molecular machinery necessary for basic functions of cellular life (4). Genes involved in DNA, RNA, protein, lipid and oligosaccharide biosynthesis as well as in energy metabolism were among those observed. Additionally, genes from other functional classes including structural proteins (*e.g.* dystroglycan and myosin light chain), signaling molecules (*e.g.* 14-3-3 proteins and MAPKK2), proteins with compartmentalized functions (*e.g.* lysosome-associated membrane glycoprotein and ER lumen retaining protein receptor 1), cell surface receptors (*e.g.* FGF receptor and STRL22 G protein coupled receptor), proteins involved in intracellular transport (*e.g.* syntaxin and alpha SNAP), membrane transporters (*e.g.* Na⁺/K⁺ ATPase and mitochondrial F1/F0 ATPase), and enzymes involved in post-translational modification and protein degradation (*e.g.* kinases, phosphatases and proteasome components) were observed and were not previously known to be ubiquitously expressed. Well known genes often used as experimental controls such glyceraldehyde 3-phosphate dehydrogenase, elongation factor 1 alpha, and gamma actin were observed but varied in expression as much as 6 fold among different cell types.

EXAMPLE 8*Genes involved in tumorigenesis*

Genes that are uniformly expressed in cancers but expressed at lower levels in normal tissues may turn out to be important for tumorigenesis, and demonstrate how gene expression patterns might be useful in the analysis of disease states. We detected 40 genes that were expressed in all cancer tissues examined at levels 3 transcript copies per cell and whose expression was at least 2-fold higher in each cancer compared to its corresponding

WO 01/38577

PCT/US00/31922

normal tissue (Table 5). Four of these transcripts had no matches to known genes and 15 matched ESTs with no known function. Several of the highly induced transcripts provided tantalizing clues about their roles in tumorigenesis. For example, S100A4 has been thought to play a role in late stage tumorigenesis as it is overexpressed in colorectal adenocarcinomas but not adenomas (30), and its induction can promote (while its inhibition can prevent) metastasis in tumor models. Midkine, a heparin-binding growth factor has been reported to be overexpressed in certain cancers (34), to transform cells *in vitro* (35), and to promote tumor angiogenesis *in vivo*. Finally, overexpression of survivin, an IAP apoptosis inhibitor (37) has been recently shown to predict shorter survival rates in colorectal cancer patients and may carry out its antiapoptotic functions as a mitotic spindle checkpoint factor (39). The observed elevated expression of such genes in many tumor types indicates a potentially general role for these genes in tumorigenesis and suggests they may be useful as diagnostic markers or targets for therapeutic intervention.

EXAMPLE 9

Estimate of gene number

The 134,135 distinct transcripts identified in this study, corresponding to approximately 84,103 unique genes, provided an estimate of gene number substantially higher than the recent estimate (~ 65,000 genes) derived from extant EST clusters. What could account for the difference between these estimates, considering that both are derived from sequencing of transcripts from similar cell types? One explanation is that the clustering estimate is based on the number of observed EST clusters (62,236) divided by a measure of the completeness of the EST database. The latter value is calculated as the fraction of "characterized" genes in GenBank that already have EST matches (~95%). The characterized genes in GenBank have been assumed to be representative of the rest of the genes in the human genome, but our SAGE data indicated that their average expression was more than 10 fold higher than the mean levels of gene expression. Similarly, the number of ESTs that were present in clusters with characterized genes was approximately 12 fold higher than clusters composed entirely of ESTs. Such highly expressed genes would be more likely to be represented in transcript databases, thereby leading to an overestimation

WO 01/38577

PCT/US00/31922

of the completeness of the EST databases, and an underestimation of the number of unique genes. Indeed, the number of UniGene clusters continues to grow as a greater diversity of tissues is analyzed through the Cancer Genome Anatomy Project, and as of the date of submission of this manuscript already exceeds the recent EST derived estimate (71,849 gene clusters in Build 80 versus 65,538 predicted from Build 70).

Like other genome-wide analyses, studies of human transcriptomes using SAGE have several potential limitations. First, a small number of transcripts would be expected to lack the restriction enzyme site required to produce the 14 bp tags, and would therefore not be detected by our analyses (12). Second, our study was limited to the 19 tissues analyzed. Genes uniquely expressed in other tissues would not have been detected, and accordingly, genes observed to be tissue specific in our studies may turn out to be expressed in other normal or disease states. Finally, identification of genes corresponding to specific tags is mainly based on large but incomplete databases of ESTs and characterized genes. SAGE tags without matches to existing databases can directly be used to identify previously uncharacterized genes (1, 12, 40), but additional 3' EST data, as well as that of genomic regions would make gene identification more rapid.

WO 01/38577

PCT/US00/31922

REFERENCES

1. Velculescu *et al.*, *Cell* **88**, 243-251 (1997).
2. Pietu *et al.*, *Genome Res* **9** 195-209 (1999).
3. Wadman, *Nature* **398**, 177 (1999).
4. Lewin, *Gene Expression* **2**, 694-727 (1980).
5. Adams *et al.*, *Nature* **377**, 3 ff. (1995)
6. Okubo *et al.*, *DNA Res* **1**, 37-45 (1994).
7. Alwine *et al.* *Proc Natl Acad Sci USA* **74**, 5350-5354 (1977).
8. Zinn *et al.* *Cell* **34**, 865-879 (1983).
9. Veres *et al.* *Science* **237**, 415-417 (1987).
10. Hedrick *et al.* *Nature* **308**, 149-153 (1984).
11. Liang & Pardee, *Science* **257**, 967-971 (1992).
12. Velculescu *et al.* *Science* **270**, 484-487 (1995).
13. Kal *et al.*, *Mol Biol Cell* **10**, 1859-1872 (1999).
14. Basrai *et al.*, NORF5/HUG1 is a component of the MEC1 mediated checkpoint response to DNA damage and replication arrest in *S. cerevisiae*. *submitted*.
15. Fields *et al.* *Nat Genet* **7**, 345-346 (1994).
16. Antequera *et al.* *Proc Natl Acad Sci USA* **90** 11995-11999 (1993).
17. Gautheret *et al.* *Genome Res* **8**, 524-530 (1998).
18. Bouck *et al.* *Trends Genet* **15**, 159-62 (1999).
19. Bentley & Groudine, *Cell* **53**, 245-256 (1988).
20. Bishop *et al.* *Nature* **250**, 199-204 (1974).
21. Mody *et al.* *Trends Neurosci* **17**, 517-25 (1994).
22. Staugaitis *et al.* *Bioessays* **18**, 13-18 (1996).
23. Mundel *et al.*, *J Cell Biol* **139**, 193-204 (1997).
24. Wiegand *et al.* *FEBS Lett* **311**, 150-154 (1992).
25. Sowden *et al.* *Differentiation* **53**, 67-74 (1993).
26. Suh & Traber, *Mol Cell Biol* **16**, 619-625 (1996).
27. Blarzino *et al.*, *Free Radic Biol Med* **26**, 446-453 (1999).
28. Busam *et al.* *Adv Anat Pathol* **6**, 12-18 (1999).

WO 01/38577

PCT/US00/31922

29. Duncan *et al.*, *Cancer Res* **58**, 1515-1520 (1998).
30. Takenage *et al.*, *Clin Cancer Res* **3**, 2309-2316 (1997).
31. Lloyd *et al.*, *Oncogene* **17**, 465-473 (1998).
32. Maelandsmo *et al.*, *Cancer Res* **56**, 5490-5498 (1996).
33. Muramatsu & Muramatsu, *Biochem Biophys Res Commun* **177**, 652-658 (1991).
34. Tsutsui *et al.*, *Cancer Res* **53**, 1281-1285 (1993).
35. Kadomatsu *et al.*, *Br J Cancer* **75**, 354-359 (1997).
36. Choudhuri *et al.*, *Cancer Res.* **57**, 1814-1819 (1997).
37. Ambrosini *et al.*, *Nat Med* **3**, 917-921 (1997).
38. Kawasaki *et al.*, *Cancer Res* **58**, 5071-5074 (1998).
39. Li *et al.*, *Nature* **396**, 580-584 (1998).
40. Polyak *et al.*, *Nature* **389**, 300-304 (1997).
41. Zhang *et al.*, *Science* **276**, 1268-1272 (1997).
42. Boukam *et al.*, *J Cell Biol* **106**, 761-771 (1988).
43. Hibi *et al.*, *Cancer Res* **58**, 5690-5694 (1998).
44. Hermeking *et al.*, *Molecular Cell* **1**, 3-11 (1997).
45. He *et al.*, *Science* **281**, 1509-1512 (1998).
46. Hastie & Bishop, *Cell* **9**, 761-774 (1976).
47. Agrawal *et al.*, *Trends Biotechnol.* **10**, 152-158 (1992)
48. Uhlmann *et al.*, *Chem. Rev.* **90**, 543-584 (1990)
49. Uhlmann *et al.*, *Tetrahedron. Lett.* **215**, 3539-3542 (1987)
50. Brown, *Meth. Mol. Biol.* **20**, 1-8 (1994)
51. Sonveaux, *Meth. Mol. Biol.* **26**, 1-72 (1994)
52. Uhlmann *et al.*, *Chem. Rev.* **90**, 543-583 (1990)
53. White & Bancroft, *J. Biol. Chem.* **257**, 8569 (1982)
54. Sambrook *et al.*, *MOLECULAR CLONING. A LABORATORY MANUAL*, 2d ed., pages 7.53-7.57 (1989)
55. Chee *et al.*, *Science* **274**, 610-14 (1996)
56. DeRisi *et al.*, *Nat. Genet.* **14**, 457-60 (1996)
57. Schena, *Bioessays* **18**, 427-31 (1996)

WO 01/38577

PCT/US00/31922

58. Lockhart *et al.*, *Nature Biotechnology*, **14** (1996)
59. Romanczuk *et al.*, *Hum. Gene. Ther.* **10**, 2615-26
60. Lanzov, *Mol. Genet. Metab.* **68**, 276-82 (1999)
61. Lai & Lien, *Exp. Nephrol.* **7**, 11-14 (1999)

WO 01/38577

PCT/US00/31922

Table 1. Tissues and transcript tags analyzed

Normal tissues	Libraries	Total Transcripts	Unique Genes
Colon epithelium ^{1,2}	2	98,089	12,941
Keratinocytes ³	2	83,835	12,598
Breast epithelium ³	2	107,632	13,429
Lung epithelium ⁴	2	111,848	11,636
Melanocytes ³	2	110,631	14,824
Prostate ³	2	98,010	9,786
Monocytes ³	3	66,673	9,504
Kidney epithelium ³	2	103,836	15,094
Chondrocytes ³	4	88,875	11,628
Cardiomyocytes ³	4	77,374	9,449
Brain ²	3	202,448	23,580
Diseased Tissues			
Colon cancer ^{1,2,3}	22	1,004,509	56,153
Pancreatic cancer ¹	4	126,414	17,050
Breast cancer ³	5	226,630	18,685
Lung cancer ⁴	5	221,302	22,783
Melanoma ³	10	269,332	25,600
Polycystic kidney dise:	2	112,839	16,280
Hemangiopericytoma ³	5	199,985	31,351
Brain cancer ²	3	186,567	23,108
Total	84	3,496,829	84,103

1 Ref. 40, 41, 44, 45

2 Lai et al.

3 unpublished

4 Ref. 43

WO 01/38577

PCT/US00/31922

Table 2. Transcript abundance

Copies/Cell	Colon Cancer Cells	
	Unique transcripts	Mass fraction mRNA (%)
>500	61	20
Match GenBank (%)	61 (100)	
50 to 500	562	27
Match GenBank (%)	554 (99)	
5 to 50	6,358	30
Match GenBank (%)	6,023 (95)	
<=5	62,400	23
Match GenBank (%)	37,536 (60)	
Total	69,381	100
Match GenBank (%)	44,174 (64)	

WO 01/38577

PCT/US00/31922

Table 3. Tissue-specific genes

Tag sequence	SEQ ID NO:	Observed	Copies/cell	Unigene Description
Colon epithelium (1.76%)				
ATACTCCACT	1	141	431	Guanylate cyclase activator 2 (guanylin, intestinal, heat-stable)
TCAGCTGCAA	2	72	220	No match
GTCAATACCA	3	57	174	H.sapiens mRNA for GCAP-II/roguanylin precursor
CCTTCAATC	4	46	141	Carbonic anhydrase I
ACACCCATCA	5	29	89	No match
CCAAACACCA	6	28	86	No match
AATAGTTTCC	7	23	70	Pregnancy-specific beta-1 glycoprotein 6
CCAGGCGTCA	8	18	55	No match
GAACAGCTCA	9	18	55	ESTs
TACTCGGCCA	10	15	46	No match
GGGGGAGAAG	11	12	37	ESTs
AGTGGGCTCA	12	11	34	No match
GAGCAACGTG	13	11	34	No match
GATCTATCCA	14	10	31	ESTs
GAAGCCGAGA	15	9	28	No match
GCCCTCGGAG	16	9	28	ESTs
ACAAGCCTAG	17	9	28	No match
GTACAGGAA	18	9	28	No match
GCCCTCGGAG	19	9	28	Human homeobox protein Cdx2 mRNA, complete cds
CTAGGATGAT	20	9	28	ESTs
CCAACATATCG	21	8	24	No match
CTGACGGGGA	22	8	24	ESTs
GAGGGTTTAA	23	8	24	Homo sapiens C19steroid specific UDP-glucuronosyltransferase mRNA, complete cds
GGGGTCCCAT	24	8	24	No match
GCCAGGTCCAC	25	7	21	No match
AGAACACCAA	26	7	21	No match
AATCCCGCCC	27	7	21	Homo sapiens hAQP8 mRNA for aquaporin 8, complete cds
ACACTGCCTC	28	6	18	No match
AGAGTCCAGG	29	6	18	Homo sapiens cardioembryonic antigen (CGM2) mRNA, complete cds
CCAGAGGTAG	30	6	18	No match
GAGGCCCCCG	31	6	18	No match
CTGTGTGCC	32	5	15	ESTs, Weakly similar to tryptase-III [H.sapiens]
GAGAGGATGG	33	5	15	ESTs
GGCTGAACCA	34	5	15	No match
CCAAATCATT	35	5	15	No match
ACGGCTGGGC	36	5	15	No match
ACCTTCATCT	37	5	15	EST
AGGGCTTGAG	38	5	15	No match
ACCTTCATCT	39	5	15	Human rearranged metabotropic glutamate receptor type II (GLUR2) mRNA, complete cds
TCAGGCCAGA	40	5	15	No match
CTGTGTGCC	41	5	15	ESTs

WO 01/38577

PCT/US00/31922

Table 3, cont.

GGATGTCAC	42	5	15	Human RecA-like protein (hREC2) mRNA, complete cds
ATCTGGAGCA	43	5	15	Alcohol dehydrogenase 1 (class I), alpha polypeptide
GAGAGGATGG	44	5	15	INTEGRAL MEMBRANE PROTEIN E16
ATCTGGAGCA	45	5	15	Alcohol dehydrogenase 3 (class I), gamma polypeptide
GGATGTCAC	46	5	15	Polymeric immunoglobulin receptor
CACAGACACA	47	4	12	No match
TGCTCCTAAC	48	4	12	No match
TATACCCGGA	49	4	12	No match
TATCCTGATG	50	4	12	No match
GGCCCTCCCG	51	4	12	No match
GTAGCGATGG	52	4	12	Pim-1 oncogene
GCAGGTGTG	53	4	12	No match
TGGGAACCGG	54	3	9	No match
ACACCTCTCT	55	3	9	No match
GGAACACAGG	56	3	9	No match
CAGCGGCAC	57	3	9	No match
CAGGTGGTC	58	3	9	Homo sapiens hRVP1 mRNA for RVP1, complete cds
GGGATAATAA	59	3	9	No match
GTGGAAATC	60	3	9	No match
GTGTGTGAAT	61	3	9	No match
ATGTGACACT	62	3	9	No match
ATGGTGTAA	63	3	9	ESTs
TCACATTGAT	64	3	9	H.sapiens mRNA for Li-cadherin
TAACATAACA	65	3	9	No match
TGCCCGGGTC	66	3	9	No match
TAGTGGGAAA	67	3	9	No match
GCTATACGGG	68	3	9	No match
TCACACCCCA	69	3	9	ESTs
CTGCCCGAAC	70	3	9	No match
AGTCACCTCT	71	3	9	No match
TCATTGGTT	72	3	9	No match
TCCTCTCCTC	73	3	9	No match
CCTCTCGGCC	74	3	9	No match
CCACTGAAGT	75	3	9	No match
CTGGCTTGCT	76	3	9	No match
GAAACACAGAA	77	3	9	EST
AAAGCACGTC	78	3	9	No match
GAAACACAGAA	79	3	9	ESTs, Weakly similar to synapse-associated protein sap47-1 [D.melanogaster]
TGATTCCAT	80	3	9	No match
AAACAGGCAC	81	3	9	No match
CTTACAGTCC	82	3	9	No match
GAATGGACTC	83	3	9	No match
GAACCCAAAC	84	3	9	No match
GAAACACAGAA	85	3	9	ESTs

Normal Brain (1.36)

WO 01/38577

PCT/US00/31922

Table 3, cont.

ACITTTGTC	86	160	237	Glial fibrillary acidic protein
GTGCAATCC	87	79	117	ESTs
CAAAAGTTA	88	36	53	ESTs
TTAACTTAT	89	33	49	Homo sapiens neuroendocrine-specific protein A (NSP) mRNA, complete cds
CAGCCAAATG	90	29	43	ESTs
GCCTGTGGTG	91	28	41	Homo sapiens LY6H mRNA, complete cds
CTTAGGACA	92	26	39	ESTs
TTGGAGGTGA	93	22	33	ESTs
ATTCCATTTC	94	20	30	ESTs
ATTCCATTTC	95	20	30	ESTs, Highly similar to RAS-RELATED PROTEIN RAB-10 (Canis familiaris)
AGAGAGCGGA	96	19	28	Human guanine nucleotide-binding regulatory protein (Go-alpha) gene
TTCTCAATAC	97	19	28	Homo sapiens mRNA for synaptopodin
CATCTCCCA	98	19	28	No match
GTATCGATT	99	16	24	Homo sapiens GABA-B receptor mRNA, complete cds
TTGTAACAG	100	15	22	ESTs, Weakly similar to cyclin I [H.sapiens]
GCCCTGTATT	101	15	22	ESTs
CCACATTGCC	102	15	22	Homo sapiens chromosome 7q22 sequence
CAGGGCAACG	103	15	22	No match
AAAAGCAAAAT	104	15	22	Human mRNA for MOBP (myelin-associated oligodendrocytic basic protein), complete cds, clone hORRP1
ACCAATCCTA	105	14	21	Human guanine nucleotide-binding regulatory protein (Go-alpha) gene
CTGTGTGTCC	106	13	19	AXONIN-1 PRECURSOR
TCAGACAATA	107	12	18	ESTs
TGGTGAGATG	108	12	18	ESTs
ATTTTGTGTT	109	12	18	ESTs
ACATTGAGTC	110	12	18	Homo sapiens mRNA for MEGF4, partial cds
GTCAGTCTAC	111	11	16	Glutamate receptor, metabotropic 3
GTCCCACTTC	112	11	16	ESTs
GGGGCCCGAA	113	11	16	No match
TGACTCACCC	114	10	15	Homo sapiens calmodulin-stimulated phosphodiesterase PDE1B1 mRNA, complete cds
GACAGCGACA	115	10	15	No match
GGTGACATA	116	10	15	ESTs
TAGCTATAAA	117	10	15	ESTs
GGTGACATA	118	10	15	ESTs
GTTCATTTT	119	10	15	ESTs
AATAAATGCG	120	10	15	ESTs
GTTCATTTT	121	10	15	ESTs
ACACATTGTA	122	10	15	No match
TACCTATTGT	123	10	15	ESTs
TTTAGCAGAA	124	10	15	Homo sapiens cyclin E2 mRNA, complete cds
TTTAGCAGAA	125	10	15	ESTs
CAATTATATGA	126	9	13	ESTs
GTGAAGGTTT	127	9	13	Homo sapiens (nuc) mRNA, complete cds
TGGACITTTTA	128	9	13	ESTs
CGATGCCACG	129	9	13	No match
GTGAAGGTTT	130	9	13	Neuron-specific RNA recognition motifs (RRMs)-containing protein [human, hippocampus, mRNA, 1992 nt]

WO 01/38577

PCT/US00/31922

Table 3, cont.

Accession	Sequence	Length	Similarity	Notes
131	TGGACTTTTA	9	13	ESTs
132	CCCTCTGTC	9	13	No match
133	TCCATTCAAG	9	13	Human done 23586 mRNA sequence
134	CTCATGTATC	8	12	No match
135	ACGACCAAT	8	12	No match
136	TATTATCTTG	8	12	ESTs
137	ACTTTATAG	8	12	ESTs
138	ACTTTATAG	8	12	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
139	CGCAGTCCGC	8	12	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
140	TGTAGTGTC	8	12	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
141	CTGCTTAAGT	8	12	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
142	ACAAGTGGAA	8	12	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
143	AATCCCAATG	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
144	ACTATGCAATC	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
145	ACGAGTCATT	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
146	TTACATTGTA	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
147	ATGCCCCCTG	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
148	TTTTATTCAT	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
149	ACAGAGCATT	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
150	TGACCAATAG	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
151	AATCCCAATG	7	10	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE
152	CGCAACTGGG	5	18	ORPHAN RECEPTOR TR4
153	GCAACACTAA	3	11	No match
154	GTAATGGATT	3	11	No match
155	AGCAGACGTG	3	11	No match
156	GGATTCCGTC	6	17	No match
157	CGGAAGCGGG	5	14	No match
158	TGTAAGTACG	5	14	No match
159	GATCAGTCAT	4	11	No match
160	GCTCAGAGTT	4	11	No match
161	TAACCTCCCC	90	241	No match
162	AGGAACAACCT	6	16	No match
163	GGGTCCGTGG	6	16	No match
164	TAGCAAAATA	5	13	No match
165	GCTGTGCACA	4	11	No match
166	CAGAAAAATCA	4	11	No match

WO 01/38577

PCT/US00/31922

Table 3, cont.

GATTTCGCTGG		4	11	No match
167				
168	Melanocyte (0.93%)			
169	GTGCCATTCT	114	309	No match
170	GATATTGTC	40	108	ESTs
171	TGATGTTTA	39	106	ESTs
172	TCAGTGAAC	27	73	5,6-DIHYDROXYINDOLE-2-CARBOXYLIC ACID OXIDASE PRECURSOR
	CCAGTCACA	21	57	ESTs, Weakly similar to LACTOSE PERMEASE [Escherichia coli]
				ESTs, Highly similar to HIGH AFFINITY GLOBULIN GAMMA FC RECEPTOR I
173	TATGAGAACC	17	46	PRECURSOR (Homo sapiens)
174	GAGTTTAGTG	16	43	No match
175	CTCCACTCTG	15	41	No match
176	ATCCAGTGAC	14	38	No match
177	TGATCTTGAG	14	38	ESTs, Moderately similar to PAS protein 5 [H. sapiens]
178	AATGGCTGTT	12	33	Human melanoma antigen recognized by T-cells (MART-1) mRNA
179	ATACTAAAA	12	33	Human cysteine protease CPP32 isoform alpha mRNA, complete cds
180	ATACTAAAA	12	33	EST
181	GTTTATTAAA	10	27	PROTEIN-TYROSINE PHOSPHATASE ZETA PRECURSOR
182	AGAAATCAGT	9	24	No match
183	TGGATATTA	9	24	Homo sapiens clone 23785 mRNA sequence
184	AATTGAGTAG	9	24	Human DNA sequence from PAC 257A7 on chromosome 6p24. Contains two unknown genes and ESTs, STSs and a GSS
185	TGAGTGCTGC	9	24	No match
186	GCAGTACAGT	8	22	No match
187	GAATTCAGGA	7	19	Homo sapiens mRNA for KIAA0679 protein, partial cds
188	GAATTCAGGA	7	19	No match
189	GAATTCAGGA	7	19	Homo sapiens melastatin 1 (MLSN1) mRNA, complete cds
190	GTTTATCTG	7	19	No match
191	GAATTCAGGA	7	19	Homo sapiens mRNA for synaposome associated protein of 23 kilodaltons, isoform A
192	GCCCGGTAG	6	16	Msh (Drosophila) homeo box homolog 1 (formerly homeo box 7)
193	TGGGGTGTGC	6	16	Homo sapiens thyroid receptor interactor (TRIP8) mRNA, 3' end of cds
194	AAATTTTATG	5	14	Interferon regulatory factor 4
195	TCAGTGTCTG	5	14	ESTs
196	GGAGGTGAGC	5	14	ESTs
197	TTCTTCTCAA	5	14	ESTs
198	TTCTTCTCAA	5	14	ESTs
199	GGTTGTCTCT	5	14	ESTs
200	CTTTGTTTAC	5	14	ESTs, Weakly similar to line-1 protein ORF2 [H. sapiens]
201	CACTATAGAA	5	14	No match
202	TTTGGTTACA	4	11	No match
203	TCAAAACAAT	4	11	EST
204	TTTGGTTACA	4	11	Human R kappa B mRNA, complete cds
205	TATAGAGCAA	4	11	Homo sapiens clone 23688 mRNA sequence
206	TAATAACCAAG	4	11	No match
207	TCTATACTG	4	11	No match
208	GGAATACGGC	4	11	No match

WO 01/38577

PCT/US00/31922

Table 3. cont.

Prostate (0.05%)					
TGAACTGGCA	209	3	9	No match	
AATGTTGGG	210	3	9	No match	
Normal Kidney (0.27%)					
CGACAAACTA	211	4	12	No match	
GTAGCACAGA	212	4	12	No match	
ACCGTCAATC	213	4	12	No match	
TGGATCAGTC	214	4	12	Human mRNA for KIAA0259 gene, partial cds	
TGGCTCGGTC	215	4	12	EST	
GCGACTGCCA	216	4	12	No match	
GCACTAGCTG	217	3	9	No match	
CGGCGCGGTT	218	3	9	No match	
CGGCAGTCCC	219	3	9	No match	
CCCCACTGT	220	3	9	No match	
CGGCGGATGG	221	3	9	No match	
CCCCAGGCCG	222	3	9	No match	
CCCATTCGAA	223	3	9	No match	
TCAAGAGGTG	224	3	9	No match	
ATAACTGTTG	225	3	9	Human HFREP-1 mRNA for unknown protein, complete cds	

WO 01/38577

PCT/US00/31922

Table 4. Ubiquitously expressed transcripts

Tag sequence	SEQ ID NO:	Copies/cell	Range	Range/Avg	UniGene Description
CATCTAACT	266	44	22	0.91	Human mRNA for KIAA0038 gene, partial cds
GGCAAGCA	267	27	14	1.00	STERIOD HORMONE RECEPTOR ERR1
ATTCAGCACC	268	29	11	1.03	ESTs. Highly similar to signal peptidase:SUBUNIT=12kD
TTGTTATGC	269	15	6	2.1	Annexin VII (synexin)
ACAGGGTGAC	270	115	47	1.04	Homo sapiens mRNA for EDF-1 protein
GCTTCCATCT	271	39	17	1.06	H. sapiens BAT1 mRNA for nuclear RNA helicase (DEAD family)
GCTTCCATCT	272	39	17	1.06	BB1-malignant cell expression-enhanced gene/tumor progression-enhanced gene
GAGGGTGGG	273	21	9	1.08	Human DR-nm23 mRNA, complete cds
GCAGGGTGGG	274	34	15	1.10	V-akt murine thymoma viral oncogene homolog 2
AGCCCTCCCT	275	85	42	1.12	Homo sapiens autoantigen p542 mRNA, complete cds
ATGGCCATAG	276	15	5	2.2	Human mRNA for YSK1, complete cds
GTGGGTGTC	277	20	9	1.13	ESTs
TGTAGTTTGA	278	41	14	1.14	Transcription elongation factor B (SIII), polypeptide 1-like
GGGGCTGTGG	279	14	6	1.15	Human TFIIC Box B-binding subunit mRNA, complete cds
GGGGCTGTGG	280	14	6	1.15	Homo sapiens mRNA for smallest subunit of ubiquinol-cytochrome c reductase, complete cds
CACGCAATGC	281	111	53	1.17	Human homolog of Drosophila enhancer of split m9/m10 mRNA, complete cds
CTCACACATT	282	49	20	1.18	LYSOSOME-ASSOCIATED MEMBRANE GLYCOPROTEIN 1 PRECURSOR
CAATGAGGA	283	36	15	1.19	Neuroblastoma RAS viral (v-ras) oncogene homolog
TGTAAGTCTG	284	21	8	3.3	Human p62 mRNA, complete cds
ACCAAGGAGG	285	63	25	1.19	ESTs
ACCAAGGAGG	286	63	25	1.19	DNA-DIRECTED RNA POLYMERASE II 23 KD POLYPEPTIDE
ACCAAGGAGG	287	63	25	1.19	Human mRNA for transcription elongation factor S-II, NS-II-T1, complete cds
TGAGGGCAGG	288	17	7	1.20	Syntaxin 5A
TCCACGCACC	289	39	14	1.20	ESTs
TAGGCAATC	290	40	14	1.21	H. sapiens mRNA for SMT3B protein
GGTAGCCTGG	291	61	25	1.21	Damage-specific DNA binding protein 1 (127 kD)
TCAACAGCCA	292	14	6	2.3	Human translation initiation factor 3 47 kDa subunit mRNA, complete cds
CTCTGTGTGG	293	18	7	2.6	Homo sapiens EB1 mRNA, complete cds
GCTCTCTATG	294	115	51	1.23	Cytochrome c oxidase subunit IV
TGCATCTGGT	295	104	32	1.24	78 KD GLUCOSE REGULATED PROTEIN PRECURSOR
GCTCTCTATG	296	72	21	1.11	H. sapiens mRNA for rat translocin-associated protein della homolog
GAAGGCATCC	297	39	16	1.25	PROBABLE 26S PROTEASE SUBUNIT TBP-1
CCACTCCTCA	298	59	19	1.26	DEFENDER AGAINST CELL DEATH 1
GCTGTCTATCA	299	31	8	1.27	26S PROTEASE REGULATORY SUBUNIT 4
CGGCTGGTGA	300	63	24	1.28	Proteasome component C5
AAGCCAGGAC	301	65	26	1.31	Homo sapiens chromosome 19, cosmid R32469
TGAGAGGGTG	302	32	15	1.32	14-3-3 PROTEIN TAU
GCGTGATCCT	303	33	10	1.32	ALCOHOL DEHYDROGENASE
CTGCCAATT	304	51	11	1.33	COFILIN, NON-MUSCLE ISOFORM
CCAAACGTGT	305	148	56	1.33	HISTONE H3.3
GCGGAGGGC	306	45	12	1.34	ADP-RIBOSYLATION FACTOR-LIKE PROTEIN 2
GGCAGGCCT	307	70	20	1.34	ESTs

WO 01/38577

PCT/US00/31922

Table 4, cont.

308	GGCCAGCCCT	70	20	134	Phosphofructokinase (liver type)
309	TGGGCAAGC	608	189	136	Translation elongation factor 1 gamma
310	GCAAAACACG	29	12	136	Human mRNA for KIAA0002 gene, complete cds
311	ACTTACTGCG	107	33	136	Cytochrome c oxidase subunit Vb
312	GTGGGCTGT	32	11	136	ESTs
313	TGCTACTGGT	18	7	136	Surfeit 1
314	GACGACACGA	401	71	137	Ribosomal protein S28
315	CAAGTGGCAA	18	5	137	Homo sapiens Grf40 adaptor protein (Grf40) mRNA, complete cds
316	TACTCTGGC	72	18	137	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN L
317	GACTGTGCCA	75	15	137	Human cytoplasmic dynein light chain 1 (hdc1) mRNA, complete cds
318	TTCGCGGTTA	318	9	137	Homo sapiens clone 24592 mRNA sequence
319	CATTGACGGA	14	5	138	Homo sapiens Chromosome 16 BAC clone CIT987SK-A-152E5
320	CAGGAACGGG	97	28	138	DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE 2
321	AATAGGTCCA	219	64	140	Ribosomal protein S25
322	ACCTCAGGAA	67	32	141	Human high density lipoprotein binding protein (HBP) mRNA, complete cds
323	ATGACTCAAG	26	12	141	Human mRNA for protein tyrosine phosphatase (PTP-BAS, type 2), complete cds
324	ATGACTCAAG	26	12	141	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0488
325	GCCTCTGCCA	26	12	141	Human mRNA for KIAA0272 gene, partial cds
326	TGCTTGTGCC	62	25	142	ADP-ribosylation factor 1
327	GGTGGCACTC	112	41	142	Aplysia ras-related homolog 12
328	GGCTGGGGT	659	168	142	H.sapiens mRNA for ribosomal protein L29
329	GGCTGGGGT	659	188	142	Homo sapiens sperm acrosomal protein mRNA, complete cds
330	CACAAACGGT	844	252	142	40S RIBOSOMAL PROTEIN S27
331	CATTGAAGGG	37	13	142	Homo sapiens clone 24433 myelodysplasia/myeloid leukemia factor 2 mRNA, complete cds
332	GTGACTGCCA	38	15	142	DPH2L=candidate tumor suppressor gene (ovarian cancer critical region of deletion)
333	GTGACTGCCA	38	15	142	Homo sapiens clone 24722 unknown mRNA, partial cds
334	AAGACAGTGG	678	222	143	Ribosomal protein L37a
335	CTGGCTGCAA	88	24	143	Cytochrome c oxidase subunit Vb
336	ACCGGGAGGT	18	5	143	Human DNA from chromosome 19-specific cosmid R27090, genomic sequence
337	ATGGAGACTT	28	8	143	Homo sapiens cltate synthase mRNA, complete cds
338	CAGCTCATCT	40	17	144	Homo sapiens hJTB mRNA, complete cds
339	ACGTGGTGAT	52	6	144	ESTs, Highly similar to LEYDIG CELL TUMOR 10 KD PROTEIN [Rattus norvegicus]
340	GCGGTGAGGT	37	9	144	ESTs, Highly similar to LEYDIG CELL TUMOR 10 KD PROTEIN [Rattus norvegicus]
341	GTGGCACACG	105	24	144	Homo sapiens small glutamine-rich tetrapeptide repeat (TPR) containing protein
342	GTGACAAACAC	42	11	145	Eukaryotic translation initiation factor 3 (eIF-3) p36 subunit
343	CTGCTATACG	228	70	145	Voltage-dependent anion channel 1
344	ACTGGCTGCT	27	10	146	Ribosomal protein L5
345	GGAGACACGG	53	16	146	ESTs
346	GGAGACACGG	53	16	146	Human antisecretory factor-1 mRNA, complete cds
347	CTGTTGGTGA	295	86	146	Tag matches ribosomal RNA sequence
348	TCAGATCTTT	358	141	146	40S RIBOSOMAL PROTEIN S23
349	TGAATGCTG	78	37	146	Ribosomal protein S4, X-linked
350	TAAGGAGCTG	289	71	146	Homo sapiens NADH:ubiquinone dehydrogenase 51 kDa subunit (NDUFV1) mRNA, nuclear gene encoding mitochondrial protein, complete cds
351	GGCTTTGGAG	41	15	146	Ribosomal protein S26
					ESTs

WO 01/38577

PCT/US00/31922

Table 4, cont.

CGCACCATG	352	41	14	-	74	1.46	GCN5-like 1=GCN5 homolog/putative regulator of transcriptional activation (clone GCN5L1)
GGCTGGTTCC	353	443	177	-	825	1.46	Homo sapiens ribosomal protein L11 mRNA, complete cds
GGGCGTGGGG	354	62	13	-	105	1.46	ESTs
CTCAGAGGAG	355	43	10	-	73	1.47	Human ribosomal protein L23-related mRNA, complete cds
TGGTCTCT	356	1233	363	-	2177	1.47	60S RIBOSOMAL PROTEIN L41
TCCCTGGCAT	357	15	5	-	27	1.47	Heterogeneous nuclear ribonucleoprotein K
GGGGGCTGCT	358	11	6	-	23	1.47	ESTs
GGGGGCTGCT	359	11	6	-	23	1.47	Human lysyl oxidase-related protein (WS9-14) mRNA, complete cds
CCACCCGAA	360	109	14	-	174	1.48	Testis enhanced gene transcript
CTGCTAGGAA	361	21	9	-	40	1.48	H. sapiens mRNA for TRAMP protein
AACTGGGCA	362	15	7	-	29	1.48	ESTs
TGGAGTGGAG	363	134	58	-	254	1.48	Human guanylate kinase (GUK1) mRNA, complete cds
TGAAGGAGCC	364	107	33	-	191	1.48	ATP SYNTHASE LIPID-BINDING PROTEIN P2 PRECURSOR
GGGGACTGAA	365	77	24	-	138	1.48	Homo sapiens mRNA for low molecular mass ubiquitinone-binding protein, complete cds
TGCACGTTTT	366	526	196	-	979	1.49	Human mRNA for antileukoprotease (ALP) from cervix uterus
CTGGATGCCG	367	33	11	-	59	1.49	Radin blood group
CCCCCTCGTG	368	24	8	-	44	1.49	Adrenergic, beta, receptor kinase 1
ATGATGCGGT	369	41	13	-	74	1.49	Cytoplasmic antiproteinase-38 kDa intracellular serine proteinase inhibitor
ATTCTCCAGT	370	356	86	-	618	1.50	Ribosomal protein L17
CCCCAGTTGC	371	219	90	-	418	1.50	Calpain, small polypeptide
CCAAGGATTG	372	21	6	-	38	1.50	Solute carrier family 5 (sodium/glucose cotransporter), member 2
GACCGAGGTG	373	25	6	-	43	1.50	Ewing sarcoma breakpoint region 1
GACTCTCTCA	374	13	5	-	25	1.50	ESTs
GACTCTGGGA	375	21	6	-	37	1.51	ESTs, Moderately similar to T13H5.2 [C.elegans]
GACTCTGGGA	376	21	6	-	37	1.51	Actin, gamma 1
CGCGCGGTG	377	207	54	-	368	1.51	Homo sapiens Chromosome 16 BAC clone C1987/SK-A-761H5
CCAGAACAGA	378	361	119	-	666	1.52	60S RIBOSOMAL PROTEIN L30
CCAGAACAGA	379	361	119	-	666	1.52	Deoxythymidylate kinase
TGGTTTTTGG	380	26	5	-	43	1.52	Homo sapiens acyl-protein thioesterase mRNA, complete cds
TTTTTGATCA	381	38	13	-	71	1.52	ER LUMEN PROTEIN RETAINING RECEPTOR 1
GTCTCCAC	382	65	24	-	122	1.52	ESTs, Highly similar to PROTEIN TRANSPORT PROTEIN SEC81 ALPHA SUBUNIT
GACCCCTGCC	383	192	30	-	323	1.52	Human FK-506 binding protein homologue (FKBP38) mRNA, complete cds
GCCCGCCTTG	384	49	16	-	91	1.52	Homo sapiens (clone mf.18) RNA polymerase II mRNA, complete cds
GGTGCTGGAG	385	24	8	-	45	1.53	Homo sapiens mRNA for putative methyltransferase
TTACCTCCTT	386	78	21	-	141	1.53	Homo sapiens 3-phosphoglycerate dehydrogenase mRNA, complete cds
AAACCAAGGC	387	18	5	-	33	1.53	ESTs
TTCTGGCTGC	388	85	11	-	141	1.53	Ubiquinol-cytochrome c reductase core protein I
TTCTGGCTGC	389	85	11	-	141	1.53	Human BAC clone RG114A06 from Tq31
CTTCTCACCG	390	33	8	-	58	1.54	Ubiquitin-conjugating enzyme E21 (homologous to yeast UBC9)
GAGAACCGTA	391	48	13	-	87	1.54	ESTs, Moderately similar to regulatory protein
GGGACCCGTA	392	658	51	-	1076	1.56	Aldolase A
GTCAAGACCA	393	28	11	-	54	1.56	Adaptin, beta 1 (beta prime)
CTGGGCTCCC	394	42	12	-	78	1.56	60S RIBOSOMAL PROTEIN L13
CGATTCTGGA	395	27	11	-	53	1.56	H. sapiens mRNA for ras-related GTP-binding protein
CAGGAGGAGT	396	73	19	-	132	1.56	PROBABLE PROTEIN DISULFIDE ISOMERASE ER-60 PRECURSOR
CAAAATCAGG	397	44	12	-	81	1.56	Human mRNA for cyclin I, complete cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

CTGGGTTAAT	398	615	116	-	1081	1.57	40S RIBOSOMAL PROTEIN S19 Hydroxacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit ESTs
TTTTCTGCTG	399	34	6	-	60	1.57	60S RIBOSOMAL PROTEIN L13A
CCCTGGCAAT	400	30	14	-	61	1.57	Homo sapiens chromosome 19, cosmid R30783
AGGCTACGGA	401	807	199	-	1472	1.58	Homo sapiens mRNA for NORI-1, complete cds
GAGGCATCC	402	23	8	-	45	1.58	ESTs, Weakly similar to MALONYL COA-ACYL CARRIER PROTEIN TRANSACYLASE [E.coli]
CTTTGATGT	403	26	11	-	52	1.58	ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit Ribosomal protein L35a
TTGGACCTGG	404	113	29	-	206	1.58	Human mRNA for Epstein-Barr virus small RNAs (EBERs) associated protein (EAP) ESTs
TTGGACCTGG	405	113	29	-	206	1.58	ESTs, Highly similar to COATOMER ZETA SUBUNIT [Bos taurus]
GTTCTGCGCA	406	213	43	-	379	1.58	Homo sapiens mRNA for KIAA0792 protein, complete cds
GATGCTGCCA	407	154	34	-	277	1.58	Homo sapiens putative fatty acid desaturase MLD mRNA, complete cds
ACGGCTCCGA	408	27	8	-	50	1.58	Thyroid autoantigen 70kD (Ku antigen)
GAGTCAGGAG	409	29	6	-	53	1.59	Homo sapiens hD54+ins2 isoform (hD54) mRNA, complete cds
GGAGGCTGAG	410	84	37	-	171	1.59	Dodecanoyl-Coenzyme A delta isomerase (3,2 trans-enoyl-Coenzyme A isomerase) ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE ASH1 SUBUNIT [Bos taurus]
GGAGGCTGAG	411	84	37	-	171	1.59	PRECURSOR [Bos taurus]
GTGATGGTGT	412	75	24	-	143	1.59	Homo sapiens foxyoliglutamate synthetase mRNA, complete cds
TCAGATGGCG	413	45	6	-	78	1.59	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEINS C1/C2
ATCGAAAGG	414	32	9	-	56	1.59	Homo sapiens DBI-related protein mRNA, complete cds
TCGCTGGTGG	415	67	26	-	133	1.60	Homo sapiens chaperonin containing t-complex polypeptide 1, delta subunit (Cctd) mRNA, complete cds
TGCTGGTGG	416	67	26	-	133	1.60	ESTs
TCAAATGCAT	417	37	9	-	68	1.60	ESTs, Moderately similar to PEANUT PROTEIN [Drosophila melanogaster]
TCCAAAGGAAT	418	13	5	-	26	1.60	Human mRNA for KIAA0064 gene, complete cds
CCCAGGGAGA	419	49	11	-	90	1.60	ESTs, Highly similar to C10 [H.sapiens]
TGGCTGCCC	420	54	15	-	102	1.60	ESTs, Highly similar to HYPOTHETICAL 6.3 KD PROTEIN ZK652.2 IN CHROMOSOME III [Caenorhabditis elegans]
TGGCTGCCC	421	54	15	-	102	1.60	ESTs, Highly similar to thymic epithelial cell surface antigen [M.musculus]
GGCCAAAGGC	422	39	14	-	77	1.60	ESTs
GGCCTGCTGC	423	69	13	-	125	1.60	H.sapiens hnRNP-E2 mRNA
GTGAAGCTGA	424	22	7	-	41	1.61	CELLULAR NUCLEIC ACID BINDING PROTEIN
GTGAAGCTGA	425	22	7	-	41	1.61	Human arginine-rich protein (ARP) gene, complete cds
GAATGTAAG	426	50	12	-	93	1.62	Homo sapiens CAG-1 7 mRNA, complete cds
GAATGTAAG	427	50	12	-	93	1.62	Homo sapiens mRNA for EDF-1 protein
CGTGTTAATG	428	73	31	-	148	1.62	ESTs, Moderately similar to NADH-UBIQUINONE OXIDOREDUCTASE B15 SUBUNIT [Bos taurus]
AGGGATTC	429	19	9	-	40	1.62	ESTs, Weakly similar to III ALU SUBFAMILY SQ WARNING ENTRY III [H.sapiens]
AGGGATTC	430	186	23	-	326	1.63	Human ribosomal protein L10 mRNA, complete cds
CAGCTCACTG	431	35	13	-	70	1.63	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX SUBUNIT VI REQUIRING PROTEIN
GTTGGCAGT	432	48	13	-	92	1.63	Homo sapiens chromosome 19, cosmid R26445
GGAGCTCTGT	433	22	5	-	42	1.63	Homo sapiens F1F0-type ATP synthase subunit q mRNA, complete cds
TGGAAGCTGTG	434	58	18	-	114	1.63	ESTs
TCTGCTTACA	435	643	205	-	1257	1.64	ESTs
AGGGCTTCCA	436	20	5	-	37	1.64	ESTs
GAGCAACGG	437	88	27	-	171	1.64	ESTs
TGTGATCAGA							

WO 01/38577

PCT/US00/31922

Table 4, cont.

438	ACACTACGGG	37	6	-	66	1.64	ESTs, Weakly similar to putative progesterone binding protein [H.sapiens]
439	AGCCAAAAA	41	12	-	79	1.64	H.sapiens hnRNP-E2 mRNA
440	GCGGGTGTGG	16	5	-	32	1.64	Human methionine aminopeptidase mRNA, complete cds
441	TTGCTAGAGG	39	13	-	78	1.65	ESTs, Weakly similar to F35H10.6 gene product [C.elegans]
442	GGGGCTTCTG	15	6	-	30	1.65	Human mRNA for cysteine protease, complete cds
443	AACCTTGAA	45	14	-	87	1.65	Human translation initiation factor eIF3 p40 subunit mRNA, complete cds
444	GTCTGACCCC	44	8	-	90	1.65	PROTEIN PHOSPHATASE PP2A, 85 KD REGULATORY SUBUNIT, ALPHA ISOFORM
445	ATGTCAATCA	48	12	-	92	1.65	Human clathrin assembly protein 50 (AP50) mRNA, complete cds
446	TCTGTCAAGA	40	15	-	81	1.66	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (oligomycin sensitivity conferring protein)
447	GCCCCAGCGA	23	6	-	46	1.66	ESTs
448	GGCAAGCCCC	425	119	-	824	1.66	Heat shock 27kD protein 1
449	CTCATCAGCT	46	16	-	95	1.66	ADENYLYL CYCLASE-ASSOCIATED PROTEIN 1
450	CTGTTGATTG	137	49	-	276	1.66	Heterogeneous nuclear ribonucleoprotein A1
451	GCCTTTAAGG	171	27	-	312	1.66	40S RIBOSOMAL PROTEIN S20
452	GCCTGAGCCT	13	6	-	28	1.66	ESTs
453	GAGCGGGATG	57	21	-	116	1.66	Proteasome (prosome, macropain) subunit, beta type, 6
454	TTACACAGTGG	56	13	-	107	1.67	Calcinurin B
455	GCCCCGTGCCA	23	8	-	48	1.67	ESTs, Highly similar to HYPOTHETICAL 38.2 KD PROTEIN IN BEM2-SPT2
456	CCCTAGGTGG	51	14	-	96	1.67	INTERGENIC REGION [Saccharomyces cerevisiae]
457	CCCTGATTTT	33	12	-	66	1.67	Human mRNA for KIAA0315 gene, partial cds
458	GTGTTAACCA	314	73	-	599	1.67	Human p97 mRNA, complete cds
459	AGGAAGCTG	469	182	-	948	1.68	Human ribosomal protein L10 mRNA, complete cds
460	TTCTCTCTGT	31	8	-	60	1.68	ESTs, Highly similar to 60S RIBOSOMAL PROTEIN L36 [Rattus norvegicus]
461	TTACTAAATG	26	5	-	48	1.68	ADP-ribosylation factor 5
462	GGGTGTGGTG	18	5	-	36	1.68	Calnexin
463	CCACTGCAGT	14	5	-	29	1.68	ESTs
464	AGCCTGGACT	47	17	-	95	1.69	GLYCOPROTEIN HORMONES ALPHA CHAIN PRECURSOR
465	GTGGGGGTGAC	24	8	-	47	1.69	Human mRNA for Mr 110,000 antigen, complete cds
466	CACTACACGG	46	11	-	88	1.69	ESTs, Weakly similar to HYPOTHETICAL 21.5 KD PROTEIN IN SEC15-SAP4
467	CTCATAGCAG	92	31	-	167	1.69	INTERGENIC REGION [S.cerevisiae]
468	GGAAATGTACG	94	27	-	187	1.70	FK506-BINDING PROTEIN PRECURSOR
469	CTGAGGGTGG	17	8	-	36	1.70	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
470	AAGGTCGAGC	75	9	-	136	1.70	Human mitochondrial ATP synthase subunit 9, P3 gene copy, mRNA, nuclear gene
471	GAATCACTGC	18	5	-	35	1.70	encoding mitochondrial protein, complete cds
472	ACATCATCGA	374	86	-	722	1.70	ESTs
473	GAATGAGGAC	27	6	-	51	1.70	60S RIBOSOMAL PROTEIN L24
474	CCCTCGCTCAG	44	14	-	89	1.70	Homo sapiens ribosomal protein L33-like protein mRNA, complete cds
475	TCCTAGCCTG	16	5	-	33	1.70	Ribosomal protein L12
476	AGGTGCGGGG	35	5	-	64	1.71	Human mRNA for reticulocalbin, complete cds
477	CTCCAATAAA	14	7	-	31	1.71	Hydroxacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit
							Homo sapiens SPF31 (SPF31) mRNA, complete cds
							Human HASNA-1 mRNA, complete cds
							Homo sapiens clone 24775 mRNA sequence

WO 01/38577

PCT/US00/31922

Table 4, cont.

478	GGCTGGAGT	73	23	-	147	1.71	ESTs, Weakly similar to HYPOTHETICAL 9.9 KD PROTEIN B0495.6 IN CHROMOSOME II [C.elegans]
479	AATTGCAAC	21	5	-	40	1.71	Homo sapiens histone macroH2A1.2 mRNA, complete cds
480	AACGCGGCA	448	22	-	790	1.71	Macrophage migration inhibitory factor
481	GGTGATATG	21	7	-	42	1.71	Homo sapiens chromosome 9, P1 clone 11659
482	GGCAACAAA	35	8	-	88	1.71	Human (clone E5.1) RNA-binding protein mRNA, complete cds
483	GGCAACAAA	35	6	-	88	1.71	Homo sapiens importin beta subunit mRNA, complete cds
484	TTTGTGACTG	28	13	-	62	1.71	Homo sapiens phosphoprotein CBP mRNA, complete cds
485	ATGAGGCCGG	23	7	-	47	1.72	No match
							Human HS1 binding protein HAX-1 mRNA, nuclear gene encoding mitochondrial protein, complete cds
486	TCAGTTTGTC	39	15	-	81	1.72	No match
487	CCCTATTAAAG	69	10	-	129	1.72	Human mRNA for KIAA0108 gene, complete cds
488	TTTCTAGTTT	55	28	-	123	1.72	Homo sapiens clone 24684 mRNA sequence
489	GGGCCCTTCC	20	5	-	40	1.72	Fibulin 1
490	GGGCCCTTCC	20	5	-	40	1.72	Homo sapiens DNA-binding protein (CROC-1B) mRNA, complete cds
491	CCTTGGTTTT	24	6	-	47	1.72	Human ras-related C3 botulinum toxin substrate (rac) mRNA, complete cds
492	GCTAAGGAGA	81	21	-	161	1.72	Human Gps1 (GPS1) mRNA, complete cds
493	TGAGGGGTGA	27	8	-	56	1.72	Ubiquitin activating enzyme E1
494	CCAGCTGCCA	63	19	-	120	1.73	No match
495	GGGCTGTTTG	16	5	-	34	1.73	Arginyl-tRNA synthetase
496	TGGACACAAAG	18	5	-	36	1.73	ESTs, Weakly similar to PUTATIVE MITOCHONDRIAL CARRIER C16C10.1 [C.elegans]
497	TCGCCAGGAA	44	12	-	89	1.73	Human mRNA for KIAA0058 gene, complete cds
498	TGATGTTTGA	24	8	-	49	1.73	No match
499	GTGGTGCACG	82	13	-	165	1.73	ESTs, Weakly similar to NUCLEAR PROTEIN SNF7 [Saccharomyces cerevisiae]
500	GTCTGCACCT	32	8	-	64	1.73	ESTs, Weakly similar to F08G12.1 [C.elegans]
501	GATGACCCCG	32	11	-	68	1.73	Ribosomal protein L9
502	ATCAAGGGTG	269	27	-	494	1.73	Human surface antigen mRNA, complete cds
503	TCTGGTCTGG	34	12	-	72	1.74	ESTs, Weakly similar to ion channel homolog RIC [M.musculus]
504	AGGATGACCC	42	6	-	79	1.74	H-sapiens mRNA for activin beta-C chain
505	AAAGGGGCA	28	9	-	59	1.74	Eukaryotic translation initiation factor 5A
506	GGCTTTACCC	178	56	-	365	1.74	Human non-histone chromosomal protein HMG-14 mRNA, complete cds
507	GCTTTTAGA	39	10	-	78	1.74	Homo sapiens clone 638 unknown mRNA, complete sequence
508	CTCTGCTCGG	18	8	-	37	1.74	ESTs
509	GCCTGGGACT	56	28	-	130	1.74	Homo sapiens clone 23930 mRNA sequence
510	GGTAGCAGGG	26	5	-	50	1.74	Homo sapiens cofactor A protein mRNA, complete cds
511	GCCGATCCCT	31	7	-	81	1.74	Cathepsin D (lysosomal aspartyl protease)
512	GCAGCTCAGG	50	13	-	101	1.74	Vacuolar H+ ATPase proton channel subunit
513	CGCAGTGTC	118	20	-	225	1.75	No match
514	CCCCTATTAA	62	13	-	121	1.75	Homo sapiens chromosome 9, P1 clone 11659
515	TTGTAAAGG	23	8	-	47	1.75	Heme oxygenase (decycling) 2
516	CCACACCCGT	17	6	-	36	1.75	Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)
517	CCTGGAAGAG	192	80	-	396	1.75	Homo sapiens alpha SNAP mRNA, complete cds
518	TAGCGCTGA	37	7	-	72	1.75	Homo sapiens Arp2/3 protein complex subunit p20-Arc (ARC20) mRNA, complete cds
519	CCTAGGACCT	19	5	-	39	1.75	Surfeit 1
520	GTGGACCCCTG	26	9	-	54	1.75	

WO 01/38577

PCT/US00/31922

Table 4, cont.

521	GTGGACCTTG	26	9	-	54	1.75	ESTs, Weakly similar to R05G6.4 gene product [C.elegans]
522	TGGGAGCAG	32	6	-	63	1.76	isoaldehyde-IRNA synthetase
523	GTCTACCTG	23	9	-	49	1.76	ESTs
524	GTACTGTGC	114	24	-	225	1.76	Homo sapiens nuclear chloride ion channel protein (NCC27) mRNA, complete cds
525	AAGATAATGC	12	5	-	27	1.76	ESTs, Weakly similar to Yel007c-ap [S.cerevisiae]
526	AATACCTCGT	31	7	-	61	1.76	ESTs
527	ACCTTGTGCC	23	6	-	47	1.76	ESTs, Weakly similar to alpha 2,6-sialyltransferase [R.norvegicus]
528	ACCTTGTGCC	23	6	-	47	1.76	Sorbitol dehydrogenase
529	GGAGGGGGCT	88	16	-	172	1.77	LAMIN A
530	GCCTATGGTC	39	9	-	78	1.77	ESTs, Highly similar to SEX-REGULATED PROTEIN JANUS-A [Drosophila melanogaster]
531	GTGCTGAATG	459	219	-	1031	1.77	MYOSIN LIGHT CHAIN ALKALI, SMOOTH-MUSCLE ISOFORM
532	TCGTGCGAGA	37	9	-	75	1.77	ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE SUBUNIT B14.5A [Bos taurus]
533	GTGACAGAAG	178	36	-	351	1.77	Eukaryotic translation initiation factor 4A (eIF-4A) isoform 1
534	TCAACGGTGT	15	5	-	31	1.77	Homo sapiens mRNA for RanBPM, complete cds
535	GAGCCTTGGT	58	11	-	113	1.77	Protein phosphatase 1, catalytic subunit, alpha isoform
536	TACATCCGAA	19	6	-	40	1.78	ESTs
537	GTCTGTGAGA	29	12	-	64	1.78	Homo sapiens mRNA for Hrs, complete cds
538	GTTAACGTC	95	18	-	187	1.78	Homo sapiens Bruton's tyrosine kinase (BTK), alpha-D-galactosidase A (GLA), L44-like ribosomal protein (L44L), and FTP3 (FTP3) genes, complete cds
539	GTGCGTAGG	141	27	-	277	1.78	ESTs, Weakly similar to F49C12.12 [C.elegans]
540	CGGATAAGGC	17	6	-	36	1.78	ESTs
541	GTCTGGGGCT	204	49	-	413	1.78	SM22-ALPHA HOMOLOG
542	CATCCTGGTG	64	12	-	125	1.78	Human mRNA for 26S proteasome subunit p97, complete cds
543	TCACAAGCAA	142	52	-	305	1.78	H sapiens alpha NAC mRNA
544	GGCTGATGTG	73	15	-	146	1.78	Glycyl-IRNA synthetase
545	CCGCTCCGGA	1272	293	-	2584	1.78	60S RIBOSOMAL PROTEIN L13
546	TCGCGGAGAA	98	33	-	208	1.78	ESTs, Weakly similar to SEX-DETERMINING TRANSFORMER PROTEIN 1 [Caenorhabditis elegans]
547	GTGCTGGAGA	98	12	-	187	1.79	Human SnRNP core protein Sm D2 mRNA, complete cds
548	TCCTCAAGAT	26	8	-	54	1.79	Human enhancer of rudimentary homolog mRNA, complete cds
549	CAACTTAGTT	60	20	-	127	1.79	Human myosin regulatory light chain mRNA, complete cds
550	GGGCGAGCTGG	35	12	-	75	1.79	ESTs
551	TTTCAGAGAG	43	8	-	84	1.79	Human calmodulin mRNA, complete cds
552	TTTCAGAGAG	43	8	-	84	1.79	Signal recognition particle 9 kD protein
553	GAGCAGAGAG	17	6	-	36	1.79	ESTs, Highly similar to ALPHA-ADAPTIN [Mus musculus]
554	GGAAGTTTCG	35	9	-	72	1.79	ESTs, Weakly similar to similar to oxysterol-binding proteins: partial CDS [C.elegans]
555	GTGCTGCCC	34	5	-	65	1.79	Homo sapiens mRNA for putative seven transmembrane domain protein
556	GCTGGGGTGG	21	6	-	44	1.79	H sapiens mRNA for mediator of receptor-induced toxicity
557	GTCAACATCT	456	99	-	918	1.80	Ribosomal protein, large, P0
558	CAAGCAGGAC	42	8	-	64	1.80	ESTs, Weakly similar to transmembrane protein [H.sapiens]
559	TTGGCTTTTC	27	8	-	57	1.80	ESTs
560	TGGCAACCTT	36	17	-	85	1.80	ESTs, Highly similar to GLUTATHIONE S-TRANSFERASE, MITOCHONDRIAL [Rattus norvegicus]
561	GCATAATAGG	391	83	-	786	1.80	Ribosomal protein L21
562	GGGGGTAAC	43	9	-	86	1.80	RNA-BINDING PROTEIN FUS/TL5

WO 01/38577

PCT/US00/31922

Table 4, cont.

563	CCTTCGAGAT	274	55	-	549	1.80	Ribosomal protein S5
564	CGGCCGCTGC	18	6	-	39	1.80	H sapiens mRNA for Glyoxalase II
565	GTGTTGCACA	210	42	-	421	1.80	Ribosomal protein S13
566	CCTCGGAAAA	156	27	-	312	1.81	60S RIBOSOMAL PROTEIN L38
567	AATAAAGGCT	56	9	-	110	1.81	Myosin, light polypeptide 3, alkali; ventricular, skeletal, slow
568	AATAAAGGCT	56	9	-	110	1.81	Aplysia ras-related homolog 9
569	CTTCTGTGTA	21	9	-	47	1.81	Homo sapiens immunophilin homolog ARA9 mRNA, complete cds
570	CTTCTGTGTA	21	9	-	47	1.81	Human mRNA for KIAA0190 gene, partial cds
571	GGTCAGTGT	144	26	-	286	1.81	Phosphoglycerate mutase 1 (brain)
572	AGCACTCCA	701	197	-	1467	1.81	Eukaryotic translation elongation factor 2
573	AAGCTGAGTG	39	12	-	62	1.81	Human M4 protein mRNA, complete cds
574	GTTCTTCGC	27	11	-	60	1.81	ESTs
575	TGAGGAATA	191	51	-	397	1.82	Triosephosphate isomerase 1
576	AGCTGTCCT	447	150	-	962	1.82	60S RIBOSOMAL PROTEIN L23
577	TACGTTCGAG	18	6	-	40	1.82	Homo sapiens GC20 protein mRNA, complete cds
578	GGGTGTGTAT	16	8	-	35	1.82	Homo sapiens angio-associated migratory cell protein (AAMP) mRNA, complete cds
579	GGAGGGATCA	37	12	-	79	1.82	Homo sapiens Integrin-linked kinase (ILK) mRNA, complete cds
580	ATCAGTGGCT	64	25	-	143	1.82	PROTEASOME BETA CHAIN PRECURSOR
581	CCCCCTGCC	57	17	-	121	1.83	ESTs
582	CCCCCTGCC	57	17	-	121	1.83	ESTs
583	CAAAAAAAA	94	8	-	180	1.83	Cholinergic receptor, nicotinic, alpha polypeptide 3
584	ACCTGCCGAC	16	5	-	37	1.83	Homo sapiens growth suppressor related (DOC-1R) mRNA, complete cds
585	GACCAAGAAA	81	17	-	185	1.83	CYTOCHROME C OXIDASE POLYPEPTIDE VIA-LIVER PRECURSOR
586	AGCCACTGG	33	9	-	69	1.83	No match
587	TTGAGCCAGC	43	21	-	101	1.83	Human KH type splicing regulatory protein KSRP mRNA, complete cds
588	TTTCAGGGGA	51	9	-	103	1.84	ESTs, Moderately similar to N-methyl-D-aspartate receptor glutamate-binding chain [R. norvegicus]
589	TCCGCCCGCG	75	32	-	169	1.84	ESTs
590	GTGATCTCG	22	6	-	46	1.84	ESTs
591	CTGCTGAGTG	46	6	-	90	1.84	ESTs, Highly similar to HYPOTHETICAL 14.1 KD PROTEIN C31A2.02 IN CHROMOSOME I [Schizosaccharomyces pombe]
592	CTGCTTAAGG	16	6	-	36	1.84	ESTs, Highly similar to HYPOTHETICAL 68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III [Caenorhabditis elegans]
593	TGTGGCTCC	33	14	-	74	1.84	ESTs, Weakly similar to No definition line found [C.elegans]
594	CGTTTCTGA	20	6	-	43	1.84	Human protein-tyrosine phosphatase (HU-PP-1) mRNA, partial sequence
595	GGAAAAAAA	97	6	-	167	1.84	Hepatocyte growth factor (hepatoplatin A; scatter factor)
596	GGAAAAAAA	97	6	-	167	1.84	ESTs, Highly similar to ATP SYNTHASE EPSILON CHAIN, MITOCHONDRIAL PRECURSOR [Bos taurus]
597	GAGGGAGTTT	546	162	-	1172	1.84	Ribosomal protein L27a
598	GACTCACTTT	156	27	-	315	1.84	Peptidylprolyl isomerase B (cyclophilin B)
599	GAGAACGGGG	33	7	-	67	1.85	ESTs, Highly similar to CORONIN [Dictyostelium discoideum]
600	TGGCTAGTGT	57	20	-	125	1.85	Human mRNA for proteasome subunit 2, complete cds
601	CTGTCATTTC	20	5	-	42	1.85	PRE-MRNA SPLICING FACTOR SRP20
602	GTCCCTGGC	320	98	-	690	1.85	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed (fox derived)
603	GCATTAAAT	76	7	-	146	1.85	ELONGATION FACTOR 1-BETA

WO 01/38577

PCT/US00/31922

Table 4, cont.

604	ATCCACATCG	69	17	-	144	1.85	ESTs, Weakly similar to CASEIN KINASE I HOMOLOG HRR25 [Saccharomyces cerevisiae]
605	CTGCTGTGAT	29	6	-	59	1.85	Human mRNA for U1 small nuclear RNP-specific C protein
606	GTGACCTCT	116	38	-	253	1.85	CYTCHROME C OXIDASE POLYPEPTIDE VIII-LIVERHEART PRECURSOR
607	GTGGACCCCA	47	9	-	97	1.86	Human slah binding protein 1 (SlahBP1) mRNA, partial cds
608	GACTAGTGCG	18	6	-	39	1.86	ESTs
609	TTATGGGATC	247	31	-	490	1.86	GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN 12.3
610	TTTCAGATTG	29	5	-	60	1.86	Human transcriptional coactivator PC4 mRNA, complete cds
611	GTCTGAGCTC	58	14	-	122	1.86	ESTs, Weakly similar to HYPOTHETICAL 15.4 KD PROTEIN C16C10.11 IN CHROMOSOME III [C.elegans]
612	CACACAATGT	22	9	-	49	1.86	Human sapiens peroxisomal phytanoyl-CoA alpha-hydroxylase (PAHX) mRNA, complete cds
613	CACACAATGT	22	9	-	49	1.86	Cytochrome c oxidase subunit IV
614	ACCCGACCCA	26	6	-	55	1.86	H sapiens mRNA for 1-acylglycerol-3-phosphate O-acyltransferase
615	GGAGGCAGGT	31	9	-	67	1.86	Human sapiens chromosome 1p33-p34 beta-1,4-galactosyltransferase mRNA, complete cds
616	TCTCAATTCT	27	8	-	58	1.87	Cell division cycle 42 (GTP-binding protein, 25kD)
617	CTCTTCAGGA	19	6	-	40	1.87	ESTs
618	CTGGGACTGC	18	7	-	40	1.87	ESTs
619	GCCGAGCAGG	28	8	-	57	1.87	ESTs
620	GCCGAGCAGG	28	8	-	57	1.87	ESTs
621	GGCCGAGGG	44	16	-	88	1.87	ESTs
622	GGGGACGGC	42	12	-	89	1.87	ESTs, Weakly similar to Y48E1B.1 [C.elegans]
623	ACTGGGTCTA	154	29	-	317	1.87	Non-metastatic cells 2, protein (NM23B) expressed in
624	GCCGAGGAAG	778	113	-	1570	1.87	Human mRNA for ribosomal protein S12
625	CAGATCTTGG	90	14	-	182	1.88	Ubiquitin A-52 residue ribosomal protein fusion product 1
626	AGGTTTCTCT	21	6	-	45	1.88	Human sapiens mRNA for proteasome subunit p58, complete cds
627	CCGTCCAAGG	532	59	-	1058	1.88	Ribosomal protein S16
628	GTGGCGGGCG	81	21	-	174	1.88	Biliary glycoprotein
629	GTGGCGGGCG	81	21	-	174	1.88	Human sapiens malignancy-associated protein mRNA, partial cds
630	GTGGCGGGCG	81	21	-	174	1.88	Human sapiens mRNA for KIAA0565 protein, complete cds
631	GGCAAGAAAG	252	34	-	507	1.88	Ribosomal protein L27
632	TCTTTACTTG	23	6	-	49	1.88	Human sapiens Arp2/3 protein complex subunit p21-Arc (ARC21) mRNA, complete cds
633	CTCCTCACCT	255	56	-	536	1.88	60S RIBOSOMAL PROTEIN L13A
634	CTCCTCACCT	255	58	-	536	1.88	Human Bak mRNA, complete cds
635	GCCTGTATGA	392	116	-	853	1.88	Ribosomal protein S24
636	GCITTAITTG	580	147	-	1203	1.88	Human mRNA fragment encoding cytoplasmic actin. (isolated from cultured epidermal cells grown from human foreskin)
637	CTTAAGGAT	27	9	-	60	1.88	ESTs, Highly similar to transcription factor ARF6 chain B [M.musculus]
638	GGATTGGCC	656	165	-	1401	1.88	Ribosomal protein, large P2
639	GGATTGGCC	656	165	-	1401	1.88	Ribosomal protein S26
640	GGATTGGCC	656	165	-	1401	1.88	Human mRNA for PIG-B, complete cds
641	TCTCCCTCC	31	5	-	62	1.89	Human mRNA for proteasome subunit Hsc7-1, complete cds
642	GGCCCTCTGA	46	9	-	96	1.89	Human peptidyl-prolyl isomerase and essential mitotic regulator (PIN1) mRNA, complete cds
643	TGGCTGTGTG	47	6	-	97	1.89	ESTs
644	AGACCAAGT	38	8	-	79	1.89	DNAJ PROTEIN HOMOLOG 1

WO 01/38577

PCT/US00/31922

Table 4, cont.

645	ATGGCCAACT	28	12	64	1.89	ESTs
646	AGGAGCTGCT	81	12	165	1.89	ESTs
647	AGGAGCTGCT	81	12	165	1.89	Human mitochondrial NADH dehydrogenase-ubiquinone Fe-S protein 8, 23 kDa subunit precursor (NDUFS8) nuclear mRNA encoding mitochondrial protein, complete cds
648	TGTACCTGTA	245	6	473	1.90	Human alpha-tubulin mRNA, complete cds
649	GATCCCAACA	70	11	143	1.90	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide
650	GGCATCTCT	38	8	80	1.90	14-3-3 PROTEIN TAU
651	AGGTGCAGAG	26	9	56	1.90	Homo sapiens pescadillo mRNA, complete cds
652	GTGGCATCAC	32	7	68	1.90	ESTs, Weakly similar to C25A1.6 [C.elegans]
653	TGTGTTGAGA	1683	321	3487	1.90	Translation elongation factor 1-alpha-1
654	GTGAGACAAA	98	14	199	1.91	Basic transcription factor 3
655	GCAACGGGCC	54	6	108	1.91	Homo sapiens mRNA for brain acyl-CoA hydrolase, complete cds
656	GCTGGCTGGC	113	27	243	1.91	Homo sapiens chaperonin containing t-complex polypeptide 1, eta subunit (CctH) mRNA, complete cds
657	GCCAAGATGC	55	11	116	1.91	ESTs
658	GCCAAGGGGC	28	8	61	1.91	Oxoglutarate dehydrogenase (lipoamide)
659	ACGGTGATGT	37	11	61	1.91	ESTs
660	CCCATCCGAA	353	77	753	1.91	Ribosomal protein L26
661	ACAACTTAG	60	24	139	1.91	Human calmodulin mRNA, complete cds
662	GCCTCCTCC	94	23	203	1.92	ESTs
663	GTGCTTGAGA	72	10	149	1.92	LAMIN A
664	TCCAATACTG	22	5	47	1.92	Human dynamin mRNA, complete cds
665	GTGGTGCGTG	39	11	88	1.92	Homo sapiens X-ray repair cross-complementing protein 2 (XRCC2) mRNA, complete cds
666	AAGAAGCAGG	38	15	88	1.92	ESTs
667	ACTTGGAGCC	42	13	95	1.92	Homo sapiens unknown mRNA, complete cds
668	CCGTGGGTAC	88	15	185	1.92	Human calmodulin mRNA, complete cds
669	ACAGTGGGGA	85	21	148	1.92	H. sapiens mRNS for clathrin-associated protein
670	ACAACTGTG	89	22	154	1.92	Human (p23) mRNA, complete cds
671	GTCTTAACTC	23	6	50	1.93	H. sapiens mRNA for Sop2p-like protein
672	CTGTGCTCGG	34	11	77	1.93	Homo sapiens Dim1p homolog (hdim1+) mRNA, complete cds
673	GTGGCCTGCA	22	5	46	1.93	ENOYL-COA HYDRATASE, MITOCHONDRIAL PRECURSOR
674	TGGTACACGT	100	43	236	1.93	ESTs, Weakly similar to K01G5.8 [C.elegans]
675	GTACTGTATG	23	9	54	1.93	Human calmodulin mRNA, complete cds
676	GTACTGTATG	23	9	54	1.93	ESTs
677	GGCCAGGTGG	25	5	53	1.93	Homo sapiens importin beta subunit mRNA, complete cds
678	GGCCAGGTGG	25	5	53	1.93	Homo sapiens calmodulin-stimulated phosphodiesterase PDE1B1 mRNA, complete cds
679	AGGAGAGGGG	20	5	43	1.93	Metalloproteinase 1 (33 kD)
680	AGGAGAGGGG	20	5	43	1.93	Homo sapiens forkhead protein FREAC-2 mRNA, complete cds
681	AGGAGAGGGG	20	5	43	1.93	Ferritin heavy chain
682	GTGGCAGGTG	100	19	213	1.93	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE T
683	TCTTGTGCAT	143	26	302	1.93	Human mRNA for KIAA0340 gene, partial cds
684	CCACACACCG	21	6	49	1.94	L-LACTATE DEHYDROGENASE M CHAIN
685	ACAAATCCTT	45	7	95	1.94	ESTs, Highly similar to HYPOTHETICAL 43.2 KD PROTEIN C34E10.1 IN CHROMOSOME III [Caenorhabditis elegans]
686	GTGAGACCCC	45	11	98	1.94	FK506-binding protein 1 (12kD)
						No match

WO 01/38577

PCT/US00/31922

Table 4, cont.

687	AAAGCAAGA	29	10	67	1.94	Electron-transfer-flavoprotein, beta polypeptide
688	CAAGGATCTA	27	12	65	1.94	Fibroblast growth factor receptor 2
689	TGAGGCGAGG	47	15	107	1.94	High mobility group box
690	TTTTGTGTGA	16	5	37	1.94	ESTs, Weakly similar to 50S RIBOSOMAL PROTEIN L20 [E.coli]
691	ACAGTCTTGC	17	6	38	1.94	CYTCHROME P450 IVF3
692	ACAGTCTTGC	17	6	38	1.94	Human mRNA for KIAA0102 gene, complete cds
693	CCAGGCACGC	40	9	87	1.95	Human HXC-26 mRNA, complete cds
694	AGTTTCCCAA	40	21	100	1.95	Homo sapiens SUL1C sulfotransferase (SUL1C) mRNA, complete cds
695	CCAGTGGCCC	274	48	582	1.95	Ribosomal protein S9
696	GCCCCGCCCT	30	11	69	1.95	Homo sapiens chromosome 19, cosmid R32184
697	TCTCTACTAA	41	6	85	1.95	Tropomyosin 4 (fibroblast)
698	CGCCTTTTCT	32	9	71	1.95	Spectrin, beta, non-erythrocytic 1
699	TGGCCCCCGC	28	8	56	1.95	ESTs
700	TGGCCCCCGC	28	8	56	1.95	Human helix-loop-helix zipper protein mRNA
701	CTCCTGGGGC	48	6	101	1.95	ESTs
702	AAGGAGCTGG	16	5	37	1.96	ESTs, Highly similar to YME1 PROTEIN [Saccharomyces cerevisiae]
703	AAGGAGCTGG	18	5	37	1.96	ESTs
704	AAGGAGCTGG	18	5	37	1.96	Homo sapiens clone lambda MEN1 region unknown protein mRNA, complete cds
705	GGCTTTGATT	18	5	40	1.96	COATOMER BETA' SUBUNIT
706	ACTACCTTCA	27	8	61	1.96	ESTs, Weakly similar to B0334.4 [C.elegans]
707	CTGTGCATT	33	11	75	1.96	Human 54 kDa protein mRNA, complete cds
708	ACTCCAAAAA	210	40	452	1.96	Human insulinoma rig-analog mRNA encoding DNA-binding protein, complete cds
709	TCCTGCCCCA	210	40	452	1.96	H. sapiens mRNA for transmembrane protein mp24
710	TCCTGCCCCA	72	14	155	1.96	Parathymosin
711	TCCTGCCCCA	72	14	155	1.96	Homo sapiens mRNA for KIAA0511 protein, partial cds
712	AAGCTGGAGG	58	15	125	1.96	Human translation initiation factor eIF3 p66 subunit mRNA, complete cds
713	GCACAAGAAG	90	19	195	1.96	ESTs
714	GAACCCGAGG	47	11	104	1.97	ESTs, Weakly similar to HYPOTHETICAL 16.8 KD PROTEIN IN SMY2-RPS101
715	GAACCCGAGG	47	11	104	1.97	INTERGENIC REGION [S.cerevisiae]
716	GCCGCGAAGC	18	5	36	1.97	Human mRNA for KIAA0029 gene, partial cds
717	CITTCAGATG	44	12	98	1.97	H. sapiens HUNK1 mRNA
718	GGCGCGTGTG	117	30	260	1.97	Phosphofructokinase, platelet
719	GTATTCCCTT	36	8	79	1.97	Homo sapiens mRNA for smallest subunit of ubiquinol-cytochrome c reductase, complete cds
720	GTATTCCCTT	36	8	79	1.97	Homo sapiens poly(A) binding protein II (PABP2) gene, complete cds
721	GTGGCCATGG	19	6	43	1.98	ESTs, Highly similar to elastin like protein [D.melanogaster]
722	GTGGTGAGCA	33	6	72	1.98	ESTs
723	GTGGTGAGCA	33	6	72	1.98	Human nicotinic acetylcholine receptor alpha6 subunit precursor, mRNA, complete cds
724	GTGGTGAGCA	33	6	72	1.98	Homo sapiens mRNA for PBK1 protein
725	CACCTAAITG	1247	410	2884	1.98	Breast cancer 1, early onset
726	CCCTTGGTTC	18	6	41	1.98	Tag matches mitochondrial sequence
727	CCCTTAGCTT	47	21	114	1.98	Homo sapiens (clone s153) mRNA fragment
728	CAGAGACGTG	30	9	88	1.98	Human mRNA for myosin regulatory light chain
729	ATGCGTGGTA	1084	174	2287	1.98	Human dysglycan (DAG1) mRNA, complete cds
730	TCGCGCTTCT	46	14	108	1.99	40S RIBOSOMAL PROTEIN S2
731	TCGTAACGAG	23	9	54	1.99	Homo sapiens flotillin-1 mRNA, complete cds
						ESTs

WO 01/38577

PCT/US00/31922

Table 4, cont.

GCGAGGAGGC	732	176	17	-	371	1.99	60S RIBOSOMAL PROTEIN L38
GCGGGGTACC	733	59	17	-	133	1.99	Human mRNA for pM5 protein
TCCTCTCCA	734	58	12	-	126	1.99	ALPHA-ACTININ 1, CYTOSKELETAL ISOFORM
CAGTCTCTCA	735	107	16	-	229	1.99	Ribosomal protein S10
ACGCTTCCT	736	56	12	-	124	1.99	ESTs, Weakly similar to VON EBNER'S GLAND PROTEIN PRECURSOR [H.sapiens]
ACGCTTCCT	737	56	12	-	124	1.99	Signal sequence receptor, beta
738							ESTs, Highly similar to HYPOTHETICAL 13.6 KD PROTEIN IN NUP170-ILS1
739							INTERGENIC REGION [Saccharomyces cerevisiae]
ATCTTCTGG	740	80	15	-	176	2.00	Human mRNA for ATP synthase gamma-subunit (L-type), complete cds
AGCTGTCCCC	741	23	5	-	50	2.00	Tyrosine 3-monoxygenase/tryptophan 5-monoxygenase activation protein, zeta polypeptide
TCITCCAGGA	742	52	11	-	114	2.00	Tag matches mitochondrial sequence
GTGCTAGGA	743	29	9	-	67	2.00	Human ribosomal protein L10 mRNA, complete cds
TGGACCCCCC	744	26	6	-	57	2.00	ESTs
ACCTGTATCC	745	158	24	-	341	2.00	ESTs, Weakly similar to K04G2.2 [C.elegans]
ACCTGCTGGT	746	17	6	-	40	2.00	INTERFERON-INDUCIBLE PROTEIN 1-8U
AGTCGTATGT	747	39	5	-	84	2.00	Homo sapiens clone 23675 mRNA sequence
TCCTACCCA	748	71	27	-	169	2.00	ESTs, Weakly similar to weak similarity to rat TEGT protein [C.elegans]
TGATTAAAGT	749	26	6	-	58	2.00	Amyloid beta (A4) precursor-like protein 2
CAGCAGAAGC	750	191	75	-	459	2.01	HEAT SHOCK FACTOR PROTEIN 1
TCCCTATTAA	751	5970	987	-	12877	2.01	Homo sapiens 4F5rei mRNA, complete cds
GTGGAGGTGC	752	42	6	-	91	2.01	No match
AAGATCCCG	753	63	15	-	142	2.01	Human 100 kDa coactivator mRNA, complete cds
GAGCGGCCTC	754	29	9	-	66	2.01	Homo sapiens DNA sequence from cosmid ICK0721Q on chromosome 6.
AACTACATAG	755	21	9	-	50	2.02	Human ORF mRNA, complete cds
GTAAGATTTG	756	33	9	-	76	2.02	ESTs
AGCCTGCAGA	757	65	17	-	147	2.02	Human 150 kDa oxygen-regulated protein ORP150 mRNA, complete cds
GGACCACTGA	758	498	174	-	1182	2.02	Homo sapiens chromosome 19, cosmid R33729
TTCAATAAAA	759	377	51	-	813	2.02	Ribosomal protein L3
CGATGGTCCC	760	377	51	-	813	2.02	TRANSCOBALAMIN I PRECURSOR
CATTGTAA	761	55	9	-	120	2.02	Ribosomal protein, large, P1
CCTGAGCCCG	762	142	23	-	309	2.02	Human B-cell receptor associated protein (hBAP) mRNA, partial cds
TGAGGCCCTCT	763	60	14	-	135	2.03	Tag matches mitochondrial sequence
AAGAGTTACG	764	29	6	-	65	2.03	ESTs, Weakly similar to ALBUMIN B-32 PROTEIN [Zea mays]
GAATCCAACT	765	17	8	-	43	2.03	ESTs
AGGGCGCAG	766	46	6	-	100	2.03	ESTs, Highly similar to 50S RIBOSOMAL PROTEIN L2 [Bacillus stearothermophilus]
GCTTAGAAGT	767	20	6	-	87	2.03	ESTs
AAGTCATTCA	768	31	6	-	69	2.03	Human SH3-containing protein EEN mRNA, complete cds
TACCCACCCC	769	31	10	-	74	2.03	HEAT SHOCK PROTEIN HSP 90-ALPHA
TACCCACCCC	770	31	10	-	74	2.03	Homo sapiens NADH-ubiquinone oxidoreductase subunit C1-B14 mRNA, complete cds
TACCCACCCC	771	57	17	-	132	2.03	H.sapiens mRNA for proc protein
CCTAGCTGGA	772	57	17	-	132	2.03	ESTs
TCGTCITTTAT	773	511	132	-	1172	2.03	Human zinc finger protein (MAZ) mRNA
774		126	18	-	275	2.04	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A
775		70	14	-	156	2.04	40S RIBOSOMAL PROTEIN S7
							UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 11 KD PROTEIN PRECURSOR

WO 01/38577

PCT/US00/31922

Table 4, cont.

776	TAGGATGGG	88	26	207	2.04	Sodium/potassium-transporting ATPase beta-3 subunit
777	GTGCATCCG	43	16	105	2.04	Casein kinase 2, beta polypeptide
778	CAGCGCTGCA	37	11	87	2.04	Human CDC37 homolog mRNA, complete cds
779	GGGAGCCCT	55	12	125	2.04	ESTs, Highly similar to BETA-ARRESTIN 2 (Homo sapiens)
780	GGGAGCCCT	55	12	125	2.04	ESTs
781	GAAGATGTGG	58	6	125	2.04	Homo sapiens clone 23967 unknown mRNA, partial cds
782	CCTACACAG	21	9	52	2.05	ESTs, Highly similar to GOLATH PROTEIN [Drosophila melanogaster]
783	TGCTAAAAA	26	9	61	2.06	Myosin, heavy polypeptide 9, non-muscle
784	CACAGAGTCC	28	7	64	2.06	Low density lipoprotein-related protein-associated protein 1 (alpha-2-macroglobulin receptor-associated protein 1)
785	GGGCCAATAA	30	8	70	2.06	Unidentified
786	GCTGCTGGG	220	49	503	2.07	Phospholipid hydroperoxide glutathione peroxidase
787	ACTGCTTGC	52	12	118	2.07	S-ADENOSYLMETHIONINE SYNTHETASE GAMMA FORM
788	ACTGCTTGCC	52	12	118	2.07	H. sapiens mRNA for Sop2p-like protein
789	CGGTTACTGT	81	20	187	2.07	Homo sapiens NADH:ubiquinone oxidoreductase NDUF S6 subunit mRNA, nuclear gene encoding mitochondrial protein, complete cds
790	AACCCGGGAG	179	50	420	2.07	Homo sapiens KIAA0408 mRNA, complete cds
791	AACCCGGGAG	179	50	420	2.07	Cytokine receptor family II, member 4
792	AACCCGGGAG	179	50	420	2.07	H. sapiens mRNA for delta 4:3-oxosteroid 5 beta-reductase
793	ATTAACAAG	98	18	220	2.07	Guanine nucleotide binding protein (G protein), alpha stimulating activity polypeptide 1
794	TTCAGTGCCT	19	6	43	2.07	ESTs, Weakly similar to GLUCOSE-6-PHOSPHATASE [Rattus norvegicus]
795	CCGTGCTCAT	51	18	123	2.07	ESTs, Highly similar to ADIPOCYTE P27 PROTEIN [Mus musculus]
796	ATCCCTCAGT	78	24	184	2.07	Activating transcription factor 4 (tax-responsive enhancer element 867)
797	TACCATCAAT	864	194	1985	2.07	Glyceroldehyde-3-phosphate dehydrogenase
798	TGCACACAG	34	14	84	2.08	Homo sapiens signal peptidase complex 18 kDa subunit mRNA, partial cds
799	GAACCTGGG	46	9	104	2.08	ESTs
800	GCCGTGTCCG	542	60	1185	2.08	Human ribosomal protein S6 mRNA, complete cds
801	ATAGAGGCAA	28	7	65	2.08	Human mRNA for KIAA0026 gene, complete cds
802	ATGTTTATG	83	11	184	2.08	Human non-histone chromosomal protein HMG-17 mRNA, complete cds
803	TAATAAAGGT	229	46	523	2.09	40S RIBOSOMAL PROTEIN S8
804	GGGATCAAGG	26	7	61	2.09	ESTs, Weakly similar to coded for by C. elegans cDNA yk157f8.5 [C.elegans]
805	CAAGGGCTTG	28	8	66	2.09	ESTs, Highly similar to RAS-RELATED PROTEIN RAP-1B [Homo sapiens; Bos taurus]
806	TGGTGTGAG	828	147	1876	2.09	Human DNA sequence from clone 1033B10 on chromosome 6p21.2-21.31
807	GAGTGAGTGA	19	8	48	2.09	ESTs, Weakly similar to C44C1.2 gene product [C.elegans]
808	GTGGCGCACA	42	9	98	2.09	Human mRNA for KIAA0072 gene, partial cds
809	ATGATCCGGA	22	5	52	2.10	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2
810	AACCTGGGAG	108	37	263	2.10	Human DNA fragmentation factor-45 mRNA, complete cds
811	AACCTGGGAG	108	37	263	2.10	Homo sapiens mRNA for KIAA0563 protein, complete cds
812	TGCTTCATCT	53	9	120	2.10	Homo sapiens androgen receptor associated protein 24 (ARA24) mRNA, complete cds
813	ATAATTGTTT	205	37	487	2.10	Ribosomal protein S29
814	GTTACGCTGT	41	9	95	2.10	Voltage-dependent anion channel 2
815	GGGAAGTCAC	22	5	60	2.10	Human FX protein mRNA, complete cds
816	GGGTGCTTGG	26	8	63	2.10	Human mRNA for ORF, Xq terminal portion
817	CAGTTACTTA	52	11	120	2.10	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide
818	GGGAAACCCC	207	70	506	2.10	Human G protein-coupled receptor (STRL22) mRNA, complete cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

819	GCCTTCCAAT	85	11	-	191	2.11	P68 PROTEIN
820	CCCTCTGGAT	485	33	-	1056	2.11	Cell division cycle 2-like 1 (PITSLRE proteins)
821	GACCTCTGTC	21	5	-	49	2.12	Homo sapiens mRNA for kinesin-like DNA binding protein, complete cds
822	GACCTCTGTC	21	5	-	49	2.12	Human SH3 domain-containing proline-rich kinase (sprk) mRNA, complete cds
823	CAGCAGTAGC	23	6	-	55	2.12	H.sapiens mRNA for 218kD Mi-2 protein
824	TTCAATTATA	47	8	-	108	2.12	Prothymosin alpha
825	CCCCACACCTA	84	15	-	160	2.12	INTESTINAL MEMBRANE A4 PROTEIN
826	GGTGGATGTG	30	6	-	69	2.12	Homo sapiens methyl-CpG binding protein MBD3 (MBD3) mRNA, complete cds
827	TCTGGTTTGT	41	5	-	91	2.12	Homo sapiens mRNA for integral membrane protein Tmp21-1 (p23)
828	TCTGGTTTGT	41	5	-	91	2.12	THYMOSIN BETA-10
829	CGCTGTAAAT	48	8	-	111	2.13	GDC21 HOMOLOG
830	TCCTGCTGCC	45	6	-	101	2.13	ESTs
831	TCCTGCTGCC	45	6	-	101	2.13	ESTs, Weakly similar to F46F6.1 [C.elegans]
832	GTGTGGTGGT	27	6	-	64	2.13	Homo sapiens mRNA for GDP dissociation inhibitor beta
833	TGATGTCCAC	10	5	-	27	2.14	ESTs
834	CCAGGAGGAA	222	77	-	551	2.14	HEAT SHOCK COGNATE 71 KD PROTEIN
835	GTGAAGCCCC	42	9	-	99	2.14	No match
836	GGGAGCCCGG	32	7	-	75	2.15	Homo sapiens herpesvirus entry protein B (HVEB) mRNA, complete cds
837	GCCATCCCT	64	14	-	150	2.15	Tag matches mitochondrial sequence
838	CAGTTGGTGG	26	8	-	69	2.15	Homo sapiens mRNA for E1B-55kDa-associated protein
839	ATCCATCTGT	21	9	-	54	2.15	H.sapiens hnRNP-E2 mRNA
840	GCCAGGAAGC	32	6	-	75	2.15	ESTs, Weakly similar to C01A2.5 [C.elegans]
841	TCCAGCCCT	32	9	-	76	2.15	ESTs, Weakly similar to T08G11.1 [C.elegans]
842	GCCCCCACT	24	6	-	58	2.15	Human MAP kinase activated protein kinase 2 mRNA, complete cds
843	TGCTGTGGT	16	5	-	45	2.15	H.sapiens BAT1 mRNA for nuclear RNA helicase (DEAD family)
844	TCCGTACAT	258	37	-	592	2.15	No match
845	GTGGTGGCA	61	12	-	144	2.15	Cholinergic receptor, nicotinic, delta polypeptide
846	GTGGTGGCA	61	12	-	144	2.15	Isovaleryl Coenzyme A dehydrogenase
847	GTGGTGGCA	61	12	-	144	2.15	Homo sapiens Josephin MJDT mRNA, complete cds
848	CTGTTAGTGT	64	13	-	130	2.16	MALATE DEHYDROGENASE, CYTOPLASMIC
849	CTCTACCCCT	68	28	-	175	2.16	Ribonuclease/angiogenin inhibitor
850	TGCTGGTGTG	30	8	-	74	2.16	Human mRNA, clone HH109 (screened by the monoclonal antibody of insulin receptor substrate-1 (IRS-1))
851	CTAAGACTTC	1455	317	-	3462	2.16	Tag matches mitochondrial sequence
852	GGAAGGACAG	39	5	-	90	2.16	ATPase, H+ transporting, lysosomal (vacuolar proton pump) 31kD
853	GAAGTGTGTC	23	9	-	60	2.16	ESTs, Highly similar to HYPOTHETICAL 37.2 KD PROTEIN C12C2.09C IN CHROMOSOME I [Schizosaccharomyces pombe]
854	GTACCCGGAC	33	9	-	81	2.17	ESTs, Weakly similar to W08E3.1 [C.elegans]
855	CTCCCTGAT	35	10	-	86	2.17	Homo sapiens dynamin (DNM) mRNA, complete cds
856	TGATCTTCAA	19	5	-	46	2.17	CALRETICULIN PRECURSOR
857	TCATCTTCAA	19	5	-	46	2.17	ESTs
858	TCATCTTCAA	19	5	-	46	2.17	RAB6, member RAS oncogene family
859	ATGTACTCTG	38	6	-	69	2.17	IMP (inosine monophosphate) dehydrogenase 2
860	CGCCGGAACA	648	123	-	1530	2.17	Ribosomal protein L4
861	AAGGAGGGT	78	14	-	184	2.17	Human phosphotyrosine independent ligand p62 for the Lck SH2 domain mRNA, complete cds
862	GAAAAAATA	112	12	-	265	2.17	Cell division cycle 10 (homologous to CDC10 of S. cerevisiae)

Table 4, cont.

863	AAACTCTGTG	27	6	-	64	2.18	Homo sapiens p120 catenin isoform 1A (CTNND1) mRNA, alternatively spliced, complete cds
864	ACACACGCAA	22	8	-	56	2.18	ESTs
865	CGCGCGAAGT	50	7	-	116	2.18	Ribosomal protein L12
866	TGTGCTAAAT	169	45	-	415	2.18	60S RIBOSOMAL PROTEIN L34
867	CGACCGTGGC	24	6	-	57	2.18	ESTs
868	GCCTGGGCTG	44	18	-	114	2.18	ESTs
869	GCCTGGGCTG	44	18	-	114	2.18	Homo sapiens molybdopterin synthase sulfurylase (MOCS3) mRNA, complete cds
870	AAAGTCAGAA	24	12	-	65	2.19	Ubiquinol-cytochrome c reductase core protein II
871	TGGAGCGCTA	31	5	-	71	2.19	ESTs, Weakly similar to PUTATIVE MITOCHONDRIAL CARRIER C16C10.1 [C.elegans]
872	GAATGATGA	70	14	-	167	2.19	Homo sapiens mRNA for c-myc binding protein, complete cds
873	TGTCGCTGGG	73	14	-	173	2.19	C4/C2 activating component of Rα-reactive factor
874	GCCCCTGCCT	39	6	-	91	2.19	Homo sapiens DNA-binding protein (CROC-1B) mRNA, complete cds
875	GCCCCTGCCT	39	6	-	91	2.19	Glutathione S-transferase M4
876	CAGGCGCTGGC	20	7	-	50	2.19	ESTs
877	CAGGCGTGGC	20	7	-	50	2.19	ESTs
878	GCAAAAAAAA	153	35	-	371	2.20	No match
879	AGCCACACAG	33	8	-	81	2.20	Human mRNA for KIAA0149 gene, complete cds
880	GAGGAAGAAG	52	16	-	130	2.20	Homologue of mouse tumor rejection antigen gp96
881	CAGCTAGTAG	20	9	-	54	2.20	Human mRNA for KIAA0174 gene, complete cds
882	TCTCTCCCT	40	10	-	99	2.20	Human mRNA for hepatoma-derived growth factor, complete cds
883	TACATTCTGT	30	7	-	74	2.20	Myeloid cell leukemia sequence 1 (BCL2-related)
884	GGGAAACCCC	39	11	-	98	2.21	ESTs, Weakly similar to HYPOTHETICAL 68.7 KD PROTEIN ZK757.1 IN CHROMOSOME III [C.elegans]
885	AGCCACTGCA	87	8	-	155	2.21	Homo sapiens mRNA for 26S proteasome subunit p55, complete cds
886	TAGTTGAAGT	55	13	-	136	2.21	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 14 KD PROTEIN
887	GCCAAAGTTG	17	5	-	43	2.21	Human mRNA for proteasome subunit p112, complete cds
888	GGCGGCTGCA	36	9	-	89	2.21	Excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)
889	AAAAAATAAA	469	38	-	1076	2.21	H.sapiens mRNA for sodium-phosphate transport system 1
890	AAAAAATAAA	469	38	-	1076	2.21	Homo sapiens GPI-linked anchor protein (GFRA1) mRNA, complete cds
891	AAAAAATAAA	469	38	-	1076	2.21	Enolase 1, (alpha)
892	AAAAAATAAA	469	38	-	1076	2.21	Calcium channel, voltage-dependent, P/Q type, alpha 1A subunit
893	TGTTCCACTC	18	5	-	46	2.21	Homo sapiens CD39L2 (CD39L2) mRNA, complete cds
894	CTCGGTGATG	30	10	-	76	2.22	H.sapiens mRNA for ras-related GTP-binding protein
895	CTTCTCAGGG	17	5	-	43	2.22	ESTs, Highly similar to PUTATIVE CYSTEINYL-TRNA SYNTHETASE C29E6.06C [Schizosaccharomyces pombe]
896	GGTAGCCAC	16	5	-	40	2.22	ESTs
897	GGGTTTTTAT	65	7	-	150	2.22	Homo sapiens dbpB-like protein mRNA, complete cds
898	CCGTGAACCC	39	12	-	99	2.23	Human translation initiation factor eIF-2alpha mRNA, 3'UTR
899	GAACAAGAT	58	5	-	133	2.23	Phosphoglycerate kinase 1
900	GATGAGTCTC	71	18	-	175	2.23	Homo sapiens proteasome subunit XAPC7 mRNA, complete cds
901	GGCCCTAGGC	43	6	-	101	2.23	H.sapiens ERF-2 mRNA
902	TGGCCCAACC	440	59	-	1041	2.23	Pyruvate kinase, muscle
903	CAGCGCGCCC	66	5	-	152	2.23	ESTs
904	AGCGGAGATC	91	27	-	231	2.24	Homo sapiens proteasome subunit XAPC7 mRNA, complete cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

905	GCGGGGTGGA	64	12	-	155	2.24	H.sapiens ERF-1 mRNA 3' end
906	GGGGCCCCCT	21	6	-	54	2.24	Homo sapiens mRNA for NA14 protein
907	AAGGAACITG	24	8	-	61	2.24	ESTs
908	AAGGAACITG	24	8	-	61	2.24	Homo sapiens clone 24655 mRNA sequence
909	AATGCAAGC	18	5	-	47	2.24	COFILIN, NON-MUSCLE ISOFORM
910	CCGTGATCC	66	22	-	171	2.25	No match
911	CCCCGCCAAG	66	11	-	159	2.25	Human adult heart mRNA for neutral calponin, complete cds
912	CTCAACAGCA	60	12	-	147	2.25	Human translation initiation factor 3 47 kDa subunit mRNA, complete cds
913	AAGGTAGCAG	58	17	-	143	2.25	ADENYL CYCLASE-ASSOCIATED PROTEIN 1
914	AAGCCAGCCG	78	5	-	180	2.25	Protein kinase C substrate 80K-H
915	CAGCCTTGA	21	5	-	52	2.25	ESTs, Weakly similar to siah binding protein 1 [H.sapiens]
916	TTTGCTCTCC	24	8	-	61	2.25	Vinculin
917	CAACATTCCT	41	14	-	106	2.26	Dopachrome tautomerase (dopachrome delta-isomerase, tyrosine-related protein 2)
918	TACTAGTCCT	77	13	-	187	2.26	HEAT SHOCK PROTEIN HSP 90-ALPHA
919	GACTCTGGTG	59	6	-	139	2.26	Homo sapiens chromosome 19, cosmid R29381
920	GACTCTGGTG	59	6	-	139	2.26	40S RIBOSOMAL PROTEIN S15A
921	GTGGCTCAGG	102	16	-	248	2.26	Homo sapiens KIAA0414 mRNA, partial cds
922	GTGGCTCAGG	102	16	-	248	2.26	Human Tax1 binding protein mRNA, partial cds
923	GTGGCGGGCA	71	16	-	177	2.27	H.sapiens mRNA for urea transporter
924	GTGGCGGGCA	71	16	-	177	2.27	Homo sapiens mRNA for KIAA0472 protein, partial cds
925	CCTGTGGTCC	66	18	-	215	2.27	No match
926	TACAGCACGG	27	6	-	68	2.27	Homo sapiens microsomal glutathione S-transferase 3 (MGST3) mRNA, complete cds
927	GTGGCACCTG	20	5	-	51	2.27	ESTs, Highly similar to NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG
928	TACACGTGAG	40	14	-	103	2.27	PRECURSOR [Xenopus laevis]
929	TCAGGCATT	69	24	-	180	2.27	ESTs, Weakly similar to GOLIATH PROTEIN [Drosophila melanogaster]
930	TTACAAAGG	26	7	-	83	2.27	ESTs, Highly similar to RAS-RELATED PROTEIN RAB-1A [H.sapiens]
931	TTCTTGTGGC	245	54	-	610	2.27	PROTEASOME ZETA CHAIN
932	TCCCTATTAG	91	14	-	220	2.27	Ribosomal protein S11
933	TACAAGAGGA	208	48	-	521	2.27	No match
934	TCAGACGCGAG	344	78	-	682	2.28	Ribosomal protein L6
935	CAGGATCCAG	35	6	-	86	2.28	Prothymosin alpha
936	TCTGTACAC	55	11	-	135	2.28	Human putative tumor suppressor (SNC6) mRNA, complete cds
937	GAAGCAGGAC	352	54	-	856	2.28	Ribosomal protein S11
938	GCGCGGCCCC	27	5	-	68	2.28	COFILIN, NON-MUSCLE ISOFORM
939	CCCTCCTGGT	69	23	-	181	2.29	ESTs, Moderately similar to nuclear autoantigen [H.sapiens]
940	TGGGCGCCTT	35	6	-	85	2.29	ESTs
941	GTGGTACAGG	121	35	-	312	2.29	Uroporphyrinogen decarboxylase
942	GGGTACAGG	121	35	-	312	2.29	Homo sapiens microtubule-based motor (HsKIFC3) mRNA, complete cds
943	GGTGAGACCT	93	43	-	255	2.29	Prostatic binding protein
944	GAGATCCGCA	59	16	-	153	2.30	INTERFERON GAMMA UP-REGULATED 1-5111 PROTEIN PRECURSOR
945	TTGGCAGCCC	48	5	-	115	2.30	Ribosomal protein L27a
946	GCCTTCCCT	22	8	-	59	2.30	APOPTOSIS REGULATOR BCL-X
947	GGAGTGGACA	190	28	-	465	2.30	60S RIBOSOMAL PROTEIN L18
948	TTATGGGGAG	28	6	-	74	2.30	H factor (complement)-like 1
949	TTATGGGGAG	28	6	-	74	2.30	TRANSFORMATION-SENSITIVE PROTEIN IEF SSP 3521

WO 01/38577

PCT/US00/31922

Table 4, cont.

950	GAATGGGGGC	43	9	-	106	2.30	ESTs, Highly similar to LYOSOMAL PRO-X CARBOXYPEPTIDASE PRECURSOR [Homo sapiens]
951	GTGGCAGGTG	192	36	-	479	2.30	No match
952	CTGGCGGTGT	126	41	-	331	2.31	ESTs
953	TGGGGTTTC	1243	255	-	3123	2.31	Ferritin heavy chain
954	GGCTGGGCGT	93	14	-	228	2.31	Clathrin, light polypeptide (Lcb)
955	GGCTGGGCGT	93	14	-	228	2.31	EST
956	CGTGTCTCC	28	6	-	73	2.31	ESTs
957	GTGTCTCATC	26	6	-	67	2.31	ESTs
958	GTGTCTCATC	26	6	-	67	2.31	Enolase 1, (alpha)
959	ACGATTGATG	23	6	-	60	2.31	ESTs, Highly similar to HYPOTHETICAL 27.5 KD PROTEIN IN SPX19-GCR2 INTERGENIC REGION [Saccharomyces cerevisiae]
960	TGTGTGTGA	75	20	-	194	2.31	Calmodulin 1 (phosphorylase kinase, delta)
961	TGGCTCCCG	49	9	-	122	2.32	H.sapiens mRNA for rho GDP-dissociation Inhibitor 1
962	ATCGGGCCCG	51	19	-	136	2.32	ESTs, Weakly similar to zinc finger protein [H.sapiens]
963	GCGGCCATCA	45	8	-	111	2.33	Human protein disulfide isomerase-related protein P5 mRNA, partial cds
964	GTGCTGGACC	63	15	-	182	2.33	Human mRNA for proteasome activator hPA28 subunit beta, complete cds
965	TGTAATCGT	206	59	-	540	2.33	Human mRNA for ornithine decarboxylase antizyme, ORF 1 and ORF 2
966	TAATGGTAAC	30	5	-	75	2.33	Homo sapiens nuclear-encoded mitochondrial cytochrome c oxidase Va subunit mRNA, complete cds
967	AACGACCTGG	156	6	-	369	2.33	Homo sapiens clone 24703 beta-tubulin mRNA, complete cds
968	GCCTGCACCC	18	7	-	49	2.34	Human neuronal octadecanoin-related ER localized protein mRNA, partial cds
969	GCCTGCACCC	18	7	-	49	2.34	ESTs
970	AAGGTGGAGG	809	156	-	2051	2.34	60S RIBOSOMAL PROTEIN L18A
971	AAGGAGATGG	487	132	-	1226	2.34	Ribosomal protein L31
972	CAGTTCTCTG	41	9	-	105	2.34	Human BTK region clone flp-3 mRNA
973	GTGAACCTC	111	38	-	287	2.35	Homo sapiens intrinsic factor-B12 receptor precursor, mRNA, complete cds
974	TAGTTGTCT	548	104	-	1386	2.35	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
975	CCTGTGACAG	61	8	-	150	2.35	Human mRNA for KIAA0106 gene, complete cds
976	CTCATAAGGA	572	118	-	1463	2.35	Tag matches mitochondrial sequence
977	GGTGGCCTTG	23	8	-	81	2.35	Homo sapiens NADH:ubiquinone oxidoreductase B12 subunit mRNA, nuclear gene encoding mitochondrial protein, complete cds
978	GCTCAGCTGG	171	29	-	432	2.36	Eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)
979	GGCCTGAGC	141	14	-	348	2.36	Human RNA polymerase II subunit (hSRP10) mRNA, complete cds
980	TCTGCTAAG	53	5	-	130	2.36	High-mobility group (nonhistone chromosomal) protein 1
981	TCTGCTAAG	53	5	-	130	2.36	ESTs
982	AGCCCCACAA	18	5	-	46	2.37	ESTs
983	CTGAGTCTCC	80	9	-	198	2.37	Guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2
984	TGCTTTGGGA	53	14	-	139	2.37	ESTs, Weakly similar to No definition line found [C.elegans]
985	CGTGTCTCTC	60	7	-	149	2.37	ESTs, Moderately similar to GTP-binding protein-associated protein [M.musculus]
986	GGGAAATCG	708	98	-	1772	2.37	THYMOSIN BETA-10
987	TCTGCTCTGGG	48	15	-	130	2.37	ESTs, Weakly similar to ori, len: 159, CAI: 0.12 [S.cerevisiae]
988	CAATAAACTG	97	12	-	242	2.37	PROTEIN TRANSLATION FACTOR SUJ1 HOMOLOG
989	GAGTCTGAGG	24	9	-	86	2.37	U1 snRNP 70K protein
990	GTGGCAGGGG	87	16	-	223	2.37	Human pancreatic zymogen granule membrane protein GP-2 mRNA, complete cds
991	GTGGCAGGGG	87	16	-	223	2.37	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)
992	CGAGGGGGCCA	186	33	-	480	2.38	Human non-muscle alpha-actinin mRNA, complete cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

993	GTGGGGGAG	19	5	-	49	2.38	Human DNA sequence from cosmid F0811 on chromosome 6. Contains Daxx, BING1, Tapasin, RGL2, KE2, BING4, BING5, ESTs and CpG islands
994	GAGTGGCTAT	26	6	-	75	2.38	Homo sapiens KIAA0419 mRNA, complete cds
995	GAGTGGCTAT	26	8	-	75	2.38	Homo sapiens mRNA for GDP dissociation inhibitor beta
996	GTAGACTCAC	17	5	-	46	2.38	LARGE PROLINE-RICH PROTEIN BAT2
997	AGGGAAGAG	27	7	-	72	2.39	Human G10 homolog (edg-2) mRNA, complete cds
998	AGGGAAGAG	27	7	-	72	2.39	Homo sapiens mRNA for KIAA0632 protein, partial cds
999	CCATCGTCC	3108	714	-	8145	2.39	Tag matches mitochondrial sequence
1000	TGCCGCGAC	34	8	-	90	2.40	No match
1001	TGTCCTGGTT	150	39	-	398	2.40	CYCLIN-DEPENDENT KINASE INHIBITOR 1
1002	CTTTTGTGC	42	6	-	107	2.40	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide
1003	ATAAATTGGG	23	8	-	62	2.40	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit b, isoform 1
1004	TATCACTCG	21	6	-	57	2.40	Human male-enhanced antigen mRNA (Mea), complete cds
1005	GTGGTGGCG	81	9	-	156	2.40	No match
1006	CCACTACACT	36	6	-	98	2.41	Human TNF-related apoptosis inducing ligand TRAIL mRNA, complete cds
1007	TGACCCACAC	29	11	-	81	2.41	ESTs, Weakly similar to F25H5.h [C.elegans]
1008	TGATTTCACT	803	132	-	2064	2.41	EST
1009	TGATTTCACT	803	132	-	2064	2.41	Tag matches mitochondrial sequence
1010	GGCTCCCACT	142	36	-	241	2.41	HEAT SHOCK PROTEIN HSP 90-BETA
1011	CCTGTGTGTG	32	6	-	82	2.41	ESTs
1012	AATCCTGTGG	514	135	-	1377	2.42	Ribosomal protein L8
1013	AGGAGCAAG	43	9	-	112	2.42	Human mRNA for NADPH-flavin reductase, complete cds
1014	CCTTTGAACA	43	7	-	111	2.42	Human Chromosome 16 BAC clone C1987SK-A-61E3
1015	GTGGGGCTAG	30	8	-	81	2.42	H. sapiens mRNA for protein phosphatase 5
1016	AGGGTGAAC	29	5	-	75	2.43	Human splicing factor SRp30c mRNA, complete cds
1017	CCTCAGGATA	270	72	-	728	2.43	ESTs
1018	CCTCAGGATA	270	72	-	728	2.43	Tag matches mitochondrial sequence
1019	TTCACATAAC	55	12	-	147	2.44	Human plectin (PLEC1) mRNA, complete cds
1020	CCCCCGTGAA	86	18	-	228	2.44	Homo sapiens interleukin-1 receptor-associated kinase (IRAK) mRNA, complete cds
1021	TGTGCTCGGG	107	35	-	295	2.44	Human mRNA for KIAA0088 gene, partial cds
1022	AAGCCTTGCT	20	6	-	54	2.44	ESTs
1023	TGTTCAATCAT	40	15	-	114	2.45	ESTs, Weakly similar to neuroendocrine-specific protein C [H. sapiens]
1024	AACATAACAAA	86	24	-	234	2.45	Ubiquitin A-52 residue ribosomal protein fusion product 1
1025	GCTGTTGCGC	158	33	-	419	2.45	40S RIBOSOMAL PROTEIN S20
1026	GGATGTGAAA	45	7	-	118	2.45	Antigen Identified by monoclonal antibodies 12E7, F21 and O13
1027	ACTGGTACGT	34	8	-	90	2.45	Homo sapiens F1Fo-ATPase synthase f subunit mRNA, complete cds
1028	TGTATTCCA	16	5	-	45	2.45	H. sapiens mRNA for alpha 4 protein
1029	GGGTGGGGGC	437	48	-	1124	2.46	Human profilin mRNA, complete cds
1030	CCACTGCACT	925	181	-	2460	2.47	Thyroid autoantigen 70kD (Ku antigen)
1031	CCACTGCACT	925	181	-	2460	2.47	Enhancer of zeste (Drosophila) homolog 1
1032	CCACTGCACT	925	181	-	2460	2.47	CD19 antigen
1033	CCACTGCACT	925	181	-	2460	2.47	Human clone 23732 mRNA, partial cds
1034	CCACTGCACT	925	181	-	2460	2.47	Annexin II (lipocortin II)
1035	CCACTGCACT	925	181	-	2460	2.47	Alkaline phosphatase, placental (Regan Isozyme)
1036	CCACTGCACT	925	181	-	2460	2.47	Homo sapiens clone 24760 mRNA sequence
1037	CCACTGCACT	925	181	-	2460	2.47	Homo sapiens carbonic anhydrase precursor (CA 12) mRNA, complete cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

CCACTGCACT	1038	925	181	2460	2.47	Homo sapiens methyl-CpG binding protein MBD4 (MBD4) mRNA, complete cds
CCACTGCACT	1039	925	181	2460	2.47	Phosphodiesterase 4C, cAMP-specific (dunce (Drosophila)-homolog phosphodiesterase IE1)
CCACTGCACT	1040	925	181	2460	2.47	Human SNRPN mRNA, 3' UTR, partial sequence
CCACTGCACT	1041	925	181	2460	2.47	Homo sapiens brachyury variant A (TBX1) mRNA, complete cds
CCACTGCACT	1042	925	181	2460	2.47	H. sapiens beta glucuronidase pseudogene
CCACTGCACT	1043	925	181	2460	2.47	G PROTEIN-ACTIVATED INWARD RECTIFIER POTASSIUM CHANNEL 4
CACTTGCCT	1044	109	21	280	2.47	ESTs, Highly similar to ACETYL-COENZYME A SYNTHETASE [Escherichia coli]
						ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT [Bos taurus]
CACCTGCCCT	1045	109	21	280	2.47	Tag matches mitochondrial sequence
GCAAGCCAAC	1046	100	17	264	2.47	Homo sapiens clone 24703 beta-tubulin mRNA, complete cds
TAGATAATGG	1047	49	5	126	2.47	Tag matches mitochondrial sequence
TCGAAGCCCC	1048	251	60	882	2.47	Endolase 1, (alpha)
AGAAAAAAA	1049	115	9	294	2.48	Human mRNA for KIAA0099 gene, complete cds
AGAAAAAAA	1050	115	9	294	2.48	Eukaryotic translation initiation factor 4A (eIF-4A) isoform 1
GGCGCTCCT	1051	66	9	172	2.48	TRANSALDOLASE
GGCGCTCCT	1052	66	9	172	2.48	ESTs
TAAACTGTTT	1053	29	7	70	2.48	40S RIBOSOMAL PROTEIN S14
TAAACTGTTT	1054	29	7	79	2.48	Human mRNA for histone H1x, complete cds
GGCCTTTTT	1055	36	6	85	2.48	Homo sapiens mRNA for KIAA0529 protein, partial cds
GGCCTTTTT	1056	36	6	85	2.48	60S RIBOSOMAL PROTEIN L24
GCGACAGCTC	1057	44	5	115	2.48	Human signal-transducing guanine nucleotide-binding regulatory (G) protein beta subunit mRNA, complete cds
CCCACACTAC	1058	57	17	159	2.49	S100 calcium-binding protein A10 (amexin II ligand, calpactin I, light polypeptide (p11))
AGCAGATCAG	1059	390	65	1034	2.49	ELONGATION FACTOR TU, MITOCHONDRIAL PRECURSOR
GCATAGGCTG	1060	90	15	240	2.49	Basigin
GAGCGCGACC	1061	25	9	72	2.49	ESTs, Weakly similar to neuroendocrine-specific protein C [H. sapiens]
AAATGCCACA	1062	42	6	110	2.49	Tag matches mitochondrial sequence
AGCCCTACAA	1063	754	206	2089	2.49	Human thymosin beta-4 mRNA, complete cds
TTGTTGAAGG	1064	399	57	1053	2.50	Homo sapiens mRNA for TRIP6 (thyroid receptor interacting protein)
TTCATACACC	1065	46	9	125	2.50	Tag matches mitochondrial sequence
CGGGGCCGAG	1066	772	125	2055	2.50	Ribosomal protein L28
GCCGGGTGGG	1067	780	96	2072	2.50	Basigin
GCAGGCATCC	1068	668	128	1786	2.50	Homo sapiens mRNA for synaptogyrin 2
GCTCCAGAC	1069	53	9	142	2.50	No match
AGCCACCCGT	1070	39	8	105	2.51	Human nuclear factor NF90 mRNA, complete cds
TCAGCTGGCC	1071	16	6	47	2.51	Adenine nucleotide translocator 3 (liver)
GGGGGGCCCT	1072	22	6	82	2.52	H. sapiens mRNA for arginine methyltransferase, splice variant, 1262 bp
CGGCCCAACG	1073	59	14	161	2.52	ESTs, Weakly similar to N-methyl-D-aspartate receptor glutamate-binding chain [R. norvegicus]
TGGCCATCTG	1074	65	14	177	2.52	Homo sapiens breakpoint cluster region protein 1 (BCRG1) mRNA, complete cds
CTTCGCCCGT	1075	59	11	159	2.52	ESTs
ACTTGTCGC	1076	27	6	73	2.52	ESTs, Highly similar to Surf-4 protein [M. musculus]
AAGACTGGCT	1077	30	6	81	2.52	ESTs, Highly similar to deduced protein product shows significant homology to coactosin from Dictyostellum discoideum [H. sapiens]
AGCACATTG	1078	42	5	112	2.53	Ribosomal protein S3A
GTGAAGGCAG	1079	467	83	1265	2.53	

WO 01/38577

PCT/US00/31922

Table 4, cont.

1080	CAATAAATGT	227	43	-	620	2.54	Ribosomal protein L37
1081	GCCAGGGCGG	46	5	-	121	2.54	ESTs, Highly similar to HYPOTHETICAL 52.8 KD PROTEIN T05E11.5 IN CHROMOSOME IV [Caenorhabditis elegans]
1082	GTGTAATAAG	57	9	-	154	2.54	Heterogeneous nuclear ribonucleoprotein A2/B1
1083	TTGTGCAGTG	25	6	-	70	2.54	Collagen, type I, alpha-2
1084	TTGTGCAGTG	25	6	-	70	2.54	ESTs
1085	GTGAAACCCC	1352	514	-	3963	2.55	Myelin oligodendrocyte glycoprotein (alternative products)
1086	GTGAAACCCC	1352	514	-	3963	2.55	Dihydropyrimidine branched chain transacylase (E2 component of branched chain keto acid dehydrogenase complex)
1087	GTGAAACCCC	1352	514	-	3963	2.55	Human mRNA for platelet-activating factor acetylhydrolase 2, complete cds
1088	GTGAAACCCC	1352	514	-	3963	2.55	GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR RECEPTOR ALPHA CHAIN PRECURSOR
1089	GTGAAACCCC	1352	514	-	3963	2.55	Thymopolein
1090	GTGAAACCCC	1352	514	-	3963	2.55	Basic fibroblast growth factor (bFGF) receptor (shorter form)
1091	GTGAAACCCC	1352	514	-	3963	2.55	Homo sapiens mRNA for KIAA0794 protein, partial cds
1092	GTGAAACCCC	1352	514	-	3963	2.55	Homo sapiens RNA polymerase I subunit hRP-A39 mRNA, complete cds
1093	GTGAAACCCC	1352	514	-	3963	2.55	Homo sapiens mRNA for KIAA0701 protein, partial cds
1094	GTGAAACCCC	1352	514	-	3963	2.55	Homo sapiens mRNA for MAX.3 cell surface antigen
1095	GTGAAACCCC	1352	514	-	3963	2.55	Homo sapiens mRNA for KIAA0706 protein, complete cds
1096	GTGAAACCCC	1352	514	-	3963	2.55	Homo sapiens deoxyribonuclease II mRNA, complete cds
1097	GTGAAACCCC	1352	514	-	3963	2.55	Homo sapiens clone 24758 mRNA sequence
1098	GTGAAACCCC	1352	514	-	3963	2.55	Kangal 1 (suppression of tumorigenicity 6, prostate; CD82 antigen (R2 leukocyte antigen, antigen detected by monoclonal and antibody 1A4))
1099	GTGAAACCCC	1352	514	-	3963	2.55	Leptin (murine obesity homolog)
1100	GACACCTGCT	45	7	-	122	2.55	ESTs, Weakly similar to TIP49 [R.norvegicus]
1101	GACGTGTGGG	94	8	-	247	2.56	H2AZ histone
1102	GCAAAACCCC	182	48	-	461	2.56	Homo sapiens tumor necrosis factor superfamily member LIGHT mRNA, complete cds
1103	TACCAAGTGA	46	6	-	124	2.56	Heat shock 60 kD protein 1 (chaperonin)
1104	CCCGTCCCCA	30	11	-	90	2.58	Chromosome 22q13 BAC Clone C17987SK.384D8 complete sequence
1105	GGTGATGAGG	35	8	-	98	2.58	Homo sapiens BC-2 protein mRNA, complete cds
1106	GTGTGTAAAA	27	8	-	78	2.59	H.sapiens CDM mRNA
1107	GGCTCCTCGA	41	11	-	117	2.59	Homo sapiens tapasin (NGS-17) mRNA, complete cds
1108	AAAGAAACT	62	12	-	174	2.60	POLYADENYLATE-BINDING PROTEIN
1109	CAGCGCACAG	22	5	-	64	2.60	ESTs
1110	CTGGGAGAGG	35	11	-	102	2.60	ESTs
1111	GAAAAATGTT	340	58	-	943	2.60	Laminin receptor (2H5 epitope)
1112	ATCAGGCCCT	192	26	-	527	2.61	Tag matches mitochondrial sequence
1113	TAGCTCTATG	107	43	-	323	2.61	ATPase, Na+/K+ transporting, alpha 1 polypeptide
1114	GATTTGGGCT	21	7	-	81	2.61	Human p76 mRNA, complete cds
1115	CCCGACGTGC	59	20	-	171	2.62	ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE B9 SUBUNIT [Bos taurus]
1116	GAAATTATGA	32	7	-	89	2.62	T-COMPLEX PROTEIN 1, ALPHA SUBUNIT
1117	TAAAAAATAA	108	7	-	290	2.63	ESTs
1118	TAAAAAATAA	108	7	-	290	2.63	Ubiquitin-conjugating enzyme E2A (RAD6 homolog)
1119	TAAAAAATAA	108	7	-	290	2.63	Homo sapiens protein kinase (BUB1) mRNA, complete cds
1120	GCCGCCCTGC	71	13	-	199	2.63	Acyl-Coenzyme A dehydrogenase, very long chain
1121	TTTGGGGGCTG	78	30	-	234	2.63	Human mRNA for proton-ATPase-like protein, complete cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

GTGGCAGGCA	1122	86	18	-	245	2.63	No match
GGCTGTACCC	1123	79	18	-	225	2.63	CYSTEINE-RICH PROTEIN
AGCAGGGCTC	1124	128	17	-	363	2.63	ESTs, Highly similar to PNG gene [H.sapiens]
AAGAAGATAG	1125	152	10	-	412	2.64	60S RIBOSOMAL PROTEIN L23A
TCTGGGGACG	1126	27	7	-	78	2.64	Human translational initiation factor 2 beta subunit (eIF-2-beta) mRNA, complete cds
GCTAGGTTTA	1127	80	9	-	220	2.65	Tag matches mitochondrial sequence
TGGTGACAGT	1128	32	6	-	91	2.65	Homo sapiens histone H2A.F/Z variant (H2AV) mRNA, complete cds
TTACCATATC	1129	196	46	-	566	2.65	Human mRNA for ribosomal protein L39, complete cds
GTGGGGGGTG	1130	59	9	-	165	2.65	No match
TGGATCCTAG	1131	28	7	-	81	2.66	Homo sapiens NADH:ubiquinone oxidoreductase NDUF53 subunit mRNA, nuclear gene encoding mitochondrial protein, complete cds
GGGTTTGAAC	1132	22	7	-	64	2.66	Homo sapiens SKB11s mRNA, complete cds
AATGCAGGCA	1133	83	9	-	231	2.67	S-adenosylhomocysteine hydrolase
ACATCGTAGG	1134	30	10	-	90	2.67	ESTs
AAGCGTGCCT	1135	59	10	-	167	2.67	Human APRT gene for adenine phosphoribosyltransferase
TGGAGGTGGG	1136	20	6	-	58	2.68	ESTs
TGCCTGCTCC	1137	21	6	-	64	2.68	ESTs
CTCCAGCTA	1138	358	87	-	1050	2.69	Annexin II (lipocortin II)
GTAAGTGATC	1139	80	8	-	223	2.69	Tag matches mitochondrial sequence
GTAAGTGATC	1140	80	8	-	223	2.69	Annexin XI (58kD autoantigen)
GTGTCGCA	1141	40	6	-	112	2.70	Homo sapiens RNA polymerase II transcription factor SIII p18 subunit mRNA, complete cds
ATCCGGCGGC	1142	114	14	-	321	2.70	Cystatin C (amyloid angiopathy and cerebral hemorrhage)
TGCTGCACCC	1143	232	61	-	688	2.70	ESTs
TTCTATTAA	1144	42	7	-	121	2.72	ESTs
CAGGAGTTCA	1145	91	23	-	270	2.72	Homo sapiens Arp2/3 protein complex subunit p34-Arc (ARC34) mRNA, complete cds
GTCTGCGTGC	1146	51	5	-	143	2.72	Proleasome component C2
GAATACAGT	1147	264	50	-	769	2.72	ESTs
GAATACAGT	1148	264	50	-	769	2.72	ESTs
TGAGCCCGGC	1149	36	8	-	106	2.74	Cathepsin D (lysosomal aspartyl protease)
GTGGGTGTGTG	1150	46	6	-	134	2.74	ESTs, Highly similar to LATENT TRANSFORMING GROWTH FACTOR BETA BINDING PROTEIN 1 PRECURSOR [Rattus norvegicus]
GTGGGTGTGTG	1151	46	6	-	134	2.74	Homo sapiens NF-AT4c mRNA, complete cds
TCACCCACAC	1152	383	111	-	1167	2.76	Acid phosphatase, prostate
TCACCCACAC	1153	383	111	-	1167	2.76	Ribosomal protein L17
CTGGATCTGG	1154	65	12	-	190	2.76	ESTs, Weakly similar to III ALU SUBFAMILY J WARNING ENTRY IIIII [H.sapiens]
GAAGATGTGT	1155	95	24	-	287	2.77	Glycogen phosphorylase B (brain form)
CGGATAACCA	1156	53	6	-	153	2.78	ESTs, Highly similar to HYPOTHETICAL 6.3 KD PROTEIN ZK652.2 IN CHROMOSOME III [Caenorhabditis elegans]
TCAGAAGGTG	1157	38	5	-	111	2.78	Human cell cycle protein p38-2G4 homolog (hG4-1) mRNA, complete cds
GAGAAACCCC	1158	95	22	-	268	2.78	ESTs, Weakly similar to RNA-binding protein [H.sapiens]
GAGAAACCCC	1159	95	22	-	268	2.78	Human mRNA for KIAA0134 gene, complete cds
GAGAAACCCC	1160	95	22	-	268	2.78	H.sapiens F11 mRNA
CTGGTTAAGA	1161	32	6	-	95	2.80	Human mRNA for KIAA0159 gene, complete cds
TTGGAGATCT	1162	93	20	-	279	2.80	Human calmodulin mRNA, complete cds
GAGGTCCTGT	1163	65	12	-	193	2.81	Human NADH:ubiquinone oxidoreductase MLRQ subunit mRNA, complete cds
TTCCGCGTGC	1164	50	5	-	146	2.81	PROTEASOME IOTA CHAIN
							Homo sapiens lysyl hydroxylase isoform 3 (PLOD3) mRNA, complete cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

1165	CAGCCCAACC	84	8	-	187	2.81	Homo sapiens eukaryotic translation initiation factor 3 subunit (p42) mRNA, complete cds
1166	GTGGCTCACA	104	9	-	303	2.81	Adenosine A2b receptor
1167	TAGAAGGCA	31	6	-	92	2.82	H.sapiens ERF-2 mRNA
1168	TAATAGACAA	33	7	-	102	2.83	ESTs, Weakly similar to putative [M.musculus]
1169	GGTGAGACAC	126	25	-	389	2.83	Adenine nucleotide translocator 3 (liver)
1170	CCCATCGTCT	39	5	-	116	2.83	No match
1171	CCGATCCAGG	59	14	-	182	2.83	Human translational initiation factor 2 beta subunit (eIF-2-beta) mRNA, complete cds
1172	GAATCGGTTA	43	10	-	133	2.83	Homo sapiens NADH-ubiquinone oxidoreductase 15kDa subunit mRNA, complete cds
1173	AACCCAGGAG	110	11	-	323	2.84	No match
1174	TTTGAAGCA	33	15	-	106	2.85	Homo sapiens hepatitis B virus X interacting protein (XIP) mRNA, complete cds
1175	CACAGGCCAA	40	8	-	122	2.85	Human mRNA for KIAA0005 gene, complete cds
1176	TCAGCTTCAC	30	7	-	93	2.85	Human mRNA for KIAA0359 gene, complete cds
1177	TCAGCTTCAC	30	7	-	93	2.85	Human putative G-protein (GP-1) mRNA, complete cds
1178	GAGGGCCGGT	81	10	-	185	2.85	ESTs, Highly similar to HISTONE H2A [Carina moschata]
1179	CCCAGCCAG	320	74	-	988	2.86	Ribosomal protein S3
1180	GTGGTGGGTG	59	5	-	176	2.86	Human RACH1 (RACH1) mRNA, complete cds
1181	CTGCCAAGTT	100	27	-	314	2.87	Homo sapiens mRNA for zyxin
1182	GAGAAACCCCT	46	12	-	144	2.87	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0506
1183	GAGAAACCCCT	46	12	-	144	2.87	Vitamin D (1,25-dihydroxyvitamin D3) receptor
1184	ACTAACACCC	544	132	-	1694	2.87	Tag matches mitochondrial sequence
1185	TTTTGGGGGC	37	7	-	112	2.88	ESTs
1186	TTTTGGGGGC	37	7	-	112	2.88	Human mRNA for proton-ATPase-like protein, complete cds
1187	GTGAAACCCA	43	15	-	140	2.88	No match
1188	GCCTTCATTG	27	12	-	89	2.89	Homo sapiens clone 23867 unknown mRNA, partial cds
1189	GTGGCAGCGA	33	6	-	101	2.89	No match
1190	GGGTCAAAG	52	14	-	165	2.89	HISTONE H3.3
1191	GGGGGTCAAC	61	9	-	186	2.90	ATP SYNTHASE LIPID-BINDING PROTEIN P1 PRECURSOR
1192	GTGAAACCCCT	864	196	-	2130	2.91	Carboxypeptidase M
1193	GTGAAACCCCT	664	198	-	2130	2.91	H.sapiens mRNA for laminin
1194	GTGAAACCCCT	864	198	-	2130	2.91	GC-RICH SEQUENCE DNA-BINDING FACTOR
1195	GTGAAACCCCT	664	198	-	2130	2.91	Homo sapiens mRNA for KIAA0596 protein, partial cds
1196	GTGAAACCCCT	864	198	-	2130	2.91	Homo sapiens clone 23605 mRNA sequence
1197	GTGAAACCCCT	664	198	-	2130	2.91	Formyl peptide receptor 1
1198	AGTTGAAATT	20	6	-	64	2.91	ESTs
1199	AGATCGCTT	74	11	-	226	2.92	Homo sapiens coatomer protein (COPA) mRNA, complete cds
1200	AGGTCAAGAG	20	7	-	65	2.92	No match
1201	CTAACGAGAC	43	11	-	136	2.93	ANGIOTENSIN-CONVERTING ENZYME PRECURSOR, SOMATIC
1202	GGGATGGCAG	38	5	-	115	2.93	VALYL-TRNA SYNTHETASE
1203	AGACCCACAA	182	39	-	512	2.93	Tag matches mitochondrial sequence
1204	TGAAGAACCC	50	7	-	155	2.94	CD63 antigen (melanoma 1 antigen)
1205	TGAATAAAA	71	6	-	214	2.95	Nucleophosmin (nucleolar phosphoprotein B23, numatrin)
1206	ACTGAGGTGC	34	9	-	109	2.95	Homo sapiens FGF-1 intracellular binding protein (FIBP) mRNA, complete cds
1207	ACTCAGAAGA	50	12	-	160	2.95	ESTs, Highly similar to NADH-UBIQUINONE OXIDOREDUCTASE AGGG SUBUNIT
1208	GAACACATCC	440	113	-	1414	2.96	PRECURSOR (Bos taurus)
1209	AACTAATACT	67	8	-	203	2.96	Ribosomal protein L19
							ESTs, Weakly similar to IIII ALU SUBFAMILY J WARNING ENTRY IIII [H.sapiens]

WO 01/38577

PCT/US00/31922

Table 4, cont.

1210	AGATGTGTGG	30	6	98	2.96	Hydroxyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit
1211	GTGGTGTGCA	27	8	89	2.97	AB
1212	GGCGTCCTGG	55	9	172	2.98	Homo sapiens RNA transcript from U17 small nucleolar RNA host gene, variant U17HG-
1213	CCTGCAATCC	47	11	152	2.98	ESTs, Weakly similar to No definition line found [C.elegans]
1214	GGCTGGCCAT	57	14	184	2.99	No match
1215	GGCTGGCCAT	57	14	184	2.99	GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN 12.3
1216	GTGCCCCCTTG	134	14	415	2.99	ESTs, Moderately similar to SULFATED SURFACE GLYCOPROTEIN 185 [Volvox carteri]
1217	GTGCCCCCTTG	134	14	415	2.99	Human alpha-tubulin mRNA, 3' end
1218	GCAGCCCCAG	90	12	281	3.00	Human alpha-tubulin mRNA, complete cds
1219	TGCTATTAAAG	160	34	515	3.00	Human transcriptional corepressor hKAP1/TIF1B mRNA, complete cds
1220	ATTGTGCCAC	34	8	110	3.00	ESTs
1221	CCATTGCACT	237	58	773	3.02	No match
1222	GCACCTCAGC	38	8	122	3.02	Alaxia telangiectasia mutated (includes complementation groups A, C and D)
1223	TTGGTCAGGC	129	24	419	3.05	ESTs
1224	TTGGTCAGGC	129	24	419	3.05	Calcium modulating ligand
1225	GGGCCCGGCA	30	6	88	3.05	Human melanoma antigen recognized by T-cells (MART-1) mRNA
1226	GTGGCACACA	70	15	228	3.06	Human mRNA for KIAA0123 gene, partial cds
1227	GTGGCACACA	70	15	228	3.06	Homo sapiens AIB1 (AIBC1) mRNA, complete cds
1228	TTGGCCAGGC	346	87	1149	3.07	Homo sapiens mRNA for MEGF8, partial cds
1229	TTGGCCAGGC	346	87	1149	3.07	Human cytochrome P450-11B (h11B3) mRNA, complete cds
1230	TTGGCCAGGC	346	87	1149	3.07	Homo sapiens X-ray repair cross-complementing protein 2 (XRCC2) mRNA, complete cds
1231	TTGGCCAGGC	346	87	1149	3.07	Homo sapiens oligodendrocyte-specific protein (OSP) mRNA, complete cds
1232	TTGGCCAGGC	346	87	1149	3.07	MHC class II transactivator
1233	TTGGCCAGGC	346	87	1149	3.07	Fc fragment of IgA, receptor for
1234	TTGGCCAGGC	346	87	1149	3.07	Protein kinase, interferon-inducible double stranded RNA dependent
1235	GTCACTGGCT	20	5	68	3.08	Zinc finger protein 157 (HZF22)
1236	GGCACCCCGT	61	8	187	3.09	Homo sapiens mRNA for Ribosomal protein kinase B (RSK-B)
1237	TCCCTATAAG	107	17	347	3.09	Glucose-6-phosphate dehydrogenase
1238	CCTGTAATCC	1302	453	4484	3.10	No match
1239	CCTGTAATCC	1302	453	4484	3.10	Breast cancer 2, early onset
1240	CCTGTAATCC	1302	453	4484	3.10	Integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)
1241	CCTGTAATCC	1302	453	4484	3.10	Transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
1242	CCTGTAATCC	1302	453	4484	3.10	Homo sapiens interferon induced tetrapeptide protein IFI60 (IFI4) mRNA, complete cds
1243	CCTGTAATCC	1302	453	4484	3.10	cds
1244	CCTGTAATCC	1302	453	4484	3.10	H. sapiens RBQ-3 mRNA
1245	CCTGTAATCC	1302	453	4484	3.10	Human hVps41p (hVPS41) mRNA, complete cds
1246	CCTGTAATCC	1302	453	4484	3.10	Human TNF-alpha converting enzyme precursor, mRNA, alternatively spliced, complete cds
1247	CCTGTAATCC	1302	453	4484	3.10	cds
1248	CCTGTAATCC	1302	453	4484	3.10	Homo sapiens mRNA for KIAA0526 protein, complete cds
						Homo sapiens melastatin 1 (MLSN1) mRNA, complete cds
						Homo sapiens clone 23716 mRNA sequence
						Homo sapiens mRNA for KIAA0538 protein, partial cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

1249	CCTGTAATCC	1302	453	4484	3.10	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, E*0101/E*0102 ALPHA CHAIN PRECURSOR
1250	CCTGTAATCC	1302	453	4484	3.10	Homo sapiens decoy receptor 2 mRNA, complete cds
1251	CCTGTAATCC	1302	453	4484	3.10	CATHEPSIN S PRECURSOR
1252	CCTGTAATCC	1302	453	4484	3.10	Homo sapiens type 6 nucleoside diphosphate kinase NM23-H6 (NM23-H6) mRNA, complete cds
1253	CCTGTAATCC	1302	453	4484	3.10	5' nucleotidase (CD73)
1254	CCTGTAATCC	1302	453	4484	3.10	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0508
1255	CCTGTAATCC	1302	453	4484	3.10	H.sapiens mRNA for p85 beta subunit of phosphatidylinositol-3-kinase
1256	CCTGTAATCC	1302	453	4484	3.10	Interleukin 12 receptor, beta-2
1257	TCCCCGTACA	3918	280	12438	3.10	No match
1258	GTACACACAC	30	9	104	3.11	ESTs
1259	GTACACACAC	30	9	104	3.11	Prothymosin alpha
1260	ATGGCAAGGG	56	9	182	3.11	ESTs, Weakly similar to III ALU SUBFAMILY J WARNING ENTRY III [H.sapiens]
1261	CTGTGGCAT	111	27	372	3.11	Ribosomal protein L21
1262	GTAGCCTCAC	623	161	2105	3.12	Actin, gamma 1
1263	AGTGAAGAC	57	10	187	3.12	Tag matches mitochondrial sequence
1264	CCTGTAGTCC	231	67	791	3.13	No match
1265	TTTTCTGAAA	66	12	218	3.13	Thioredoxin
1266	CTCCCGTGCC	62	9	203	3.14	Capping protein (actin filament), gelsolin-like
1267	TCTCTTTTC	32	6	108	3.14	H.sapiens tissue specific mRNA
1268	GCGGACGAGG	35	8	118	3.14	Homo sapiens TFAR19 mRNA, complete cds
1269	GCGGACGAGG	35	8	118	3.14	Human tip associating protein (TAP) mRNA, complete cds
1270	GGAGTCATTG	56	12	190	3.16	Human mRNA for proteasome subunit Hsc10-II, complete cds
1271	GTAGCAGGTG	67	21	233	3.17	Homo sapiens cargo selection protein TIP47 (TIP47) mRNA, complete cds
1272	CGCAAGCTGG	65	13	221	3.17	LAMIN A
1273	GTAAACCCG	36	11	126	3.18	No match
1274	AGGTACGAG	359	133	1274	3.18	Major histocompatibility complex, class II, DR beta 5
1275	AGGTACGAG	359	133	1274	3.18	Human mRNA for KIAA0331 gene, complete cds
1276	AGGTACGAG	359	133	1274	3.18	Human mRNA for KIAA0226 gene, complete cds
1277	GAATGCAGTT	13	5	45	3.18	ESTs
1278	GAATGCAGTT	13	5	45	3.18	ESTs
1279	GAATGCAGTT	13	5	45	3.18	ESTs
1280	GTAGGCCAT	77	21	269	3.21	HEAT SHOCK PROTEIN HSP 90-BETA
1281	GTAATCCTGC	109	23	375	3.22	Tag matches ribosomal RNA sequence
1282	TGAAGTAACA	31	7	108	3.22	PROTEIN TRANSLATION FACTOR SUI1 HOMOLOG
1283	TGCCTGTAAT	59	15	208	3.22	ISLET AMYLOID POLYPEPTIDE PRECURSOR
1284	GTAGCATAAA	28	6	95	3.23	Human ubiquitin gene, complete cds
1285	CCGTGGTCTG	67	9	224	3.23	Fibrillarin
1286	ATGAACCCC	67	24	240	3.23	Homo sapiens mRNA expressed in osteoblast, complete cds
1287	AAGATTGGTG	81	13	275	3.25	CD9 antigen
1288	ATCCGTGCC	35	11	124	3.25	Human calmodulin mRNA, complete cds
1289	CCCTTCACTG	16	5	58	3.26	ESTs, Moderately similar to III ALU SUBFAMILY J WARNING ENTRY III [H.sapiens]
1290	CCCTTCACTG	16	5	58	3.26	ESTs
1291	CAGCTGGGGC	54	6	163	3.26	Polypyrimidine tract binding protein (hnRNP I) (alternative products)
1292	CAGGCCCCAC	109	17	370	3.26	Human mRNA for calgizzarin, complete cds
1293	TGTTTATCCT	25	7	89	3.26	.

WO 01/38577

PCT/US00/31922

Table 4, cont.

1294	TAACCAATCA	52	14	-	184	3.26	Human Rab5c-like protein mRNA, complete cds
1295	CACCTGTAGT	32	5	-	110	3.27	Ribosomal protein L5
1296	TACCCCTAAA	103	16	-	351	3.27	Human kpn1 repeat mma (cdna clone pcd-kpn1-4), 3' end
1297	TACCCCTAAA	103	16	-	351	3.27	Homo sapiens mRNA for KIAA0675 protein, complete cds
1298	TACCCCTAAA	103	16	-	351	3.27	Human Line-1 repeat mRNA with 2 open reading frames
1299	TGCTCTGTGG	175	83	-	655	3.28	Human platelet-endothelial tetraspan antigen 3 mRNA, complete cds
1300	GCAAAACCCCT	81	19	-	284	3.28	No match
1301	AAGGACCTTT	115	18	-	398	3.28	ESTs
1302	CTGGCGCCGA	39	9	-	138	3.30	ESTs, Weakly similar to F55G12.9 [C.elegans]
1303	GAAGCTTTGC	133	15	-	454	3.30	HEAT SHOCK PROTEIN HSP 90-ALPHA
1304	GGTCGAGCG	57	8	-	195	3.30	Ribosomal protein S16
1305	TTGCCAGGC	69	21	-	251	3.30	Cell division cycle 42 (GTP-binding protein, 25kD)
1306	TTGCCAGGC	69	21	-	251	3.30	Human brain mRNA homologous to 3'UTR of human CD24 gene, partial sequence
1307	ACCAACGTCA	55	9	-	189	3.31	Jun B proto-oncogene
1308	GTCCCACTGG	29	8	-	103	3.31	Mannose-6-phosphate receptor (cation dependent)
1309	TTTACGGCC	142	18	-	489	3.31	Tag matches mitochondrial sequence
1310	CTGTATATCC	71	11	-	248	3.32	ESTs, Moderately similar to III ALU SUBFAMILY J WARNING ENTRY III [H.sapiens]
1311	CAGTTTGGG	47	8	-	165	3.33	ESTs
1312	CGGGGTGATG	92	20	-	325	3.33	Human copper transport protein HAH1 (HAH1) mRNA, complete cds
1313	GGGGTAAGAA	82	6	-	213	3.33	Prostatic binding protein
1314	TGACTGGCAG	49	7	-	172	3.34	CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5, EJ16, EJ30, EL32 and G344)
1315	CAATGTGTTA	47	17	-	178	3.39	H.sapiens mRNA for NADH dehydrogenase
1316	GGCTCGGGAT	74	6	-	257	3.40	CALPAIN 1, LARGE
1317	TGCTGTAGT	71	15	-	258	3.40	Hum ORF (CE15) mRNA, 3' flank
1318	CGCCGCCGGC	807	148	-	2908	3.42	Human ribosomal protein L35 mRNA, complete cds
1319	GGTGGGGAGA	68	8	-	239	3.44	Human chromosome 17q21 mRNA clone LF113
1320	GTAACCCCT	24	8	-	90	3.44	No match
1321	GGCTCTGGC	100	9	-	354	3.44	Homo sapiens b(2)gon homolog mRNA, complete cds
1322	AGTAGGTGGC	53	5	-	188	3.46	Tag matches mitochondrial sequence
1323	GGAGGTGGGG	128	19	-	456	3.48	Granulin
1324	CCITTGGCTA	27	5	-	100	3.49	ESTs, Highly similar to 40S RIBOSOMAL PROTEIN S27 [Rattus norvegicus]
1325	AGAAAGATGT	74	11	-	268	3.50	Annexin I (lipocortin I)
1326	AGAACAAAC	75	6	-	271	3.52	Proliferation-associated gene A (natural killer-enhancing factor A)
1327	AACTAAAAA	110	9	-	396	3.53	Ubiquitin A-52 residue ribosomal protein fusion product 1
1328	ATTGCACAC	38	5	-	138	3.53	Human transglutaminase mRNA, 3' untranslated region
1329	GATCCCAACT	389	27	-	1402	3.54	H.sapiens mRNA for metallothionein isoform 2
1330	GATCCCAACT	389	27	-	1402	3.54	Human mRNA for metallothionein from cadmium-treated cells
1331	CAGTACTGAC	356	99	-	1381	3.54	Tag matches mitochondrial sequence
1332	CTGTACAGAC	132	20	-	487	3.55	Homo sapiens beta 2 gene
1333	TACCCCTAGAA	43	5	-	159	3.58	Estrogen receptor
1334	GTAACACCCC	57	6	-	213	3.58	Tumor necrosis factor receptor 2 (75kD)
1335	GTAACACCCC	57	6	-	213	3.58	Homo sapiens mRNA for KIAA0632 protein, partial cds
1336	GTAACACCCC	57	6	-	213	3.58	Homo sapiens protease-activated receptor 4 mRNA, complete cds
1337	CTGAGAGGCTG	32	9	-	125	3.61	Homo sapiens growth-arrest-specific protein (gas) mRNA, complete cds
1338	GGCTGGTCTG	57	6	-	211	3.62	ESTs
1339	ACGCAGGGAG	360	29	-	1334	3.63	HEAT SHOCK PROTEIN HSP 90-ALPHA

WO 01/38577

PCT/US00/31922

Table 4, cont.

1340	GCCCTGGGCC	44	5	-	165	3.63	Homo sapiens mRNA for protein phosphatase 2C gamma
1341	CTCCCTTGCC	20	5	-	78	3.64	ESTs, Highly similar to COATOMER ZETA SUBUNIT [Bos taurus]
1342	CTGTAACT	81	27	-	323	3.65	V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3 (alternative products)
1343	AGTCTCTAGC	391	18	-	1448	3.66	Glutathione-S-transferase pi-1
1344	ACTGAAGCG	68	15	-	268	3.68	Human melargidin precursor mRNA, complete cds
1345	AAGGAAGATG	24	6	-	94	3.68	PROTEASOME COMPONENT C13 PRECURSOR
1346	CCGACGGGGC	60	14	-	237	3.71	Tag matches ribosomal RNA sequence
1347	GCCCCCAATA	428	6	-	1601	3.73	Lectin, galactoside-binding, soluble, 1 (galecin 1)
1348	AGATGTGGG	49	9	-	193	3.74	Homo sapiens mRNA for KIAA0706 protein, complete cds
1349	GGAGGCCGAG	26	5	-	103	3.75	ESTs, Weakly similar to allograft inflammatory factor-1 [H.sapiens]
1350	ACCCGCCGCG	65	6	-	251	3.76	Jun D proto-oncogene
1351	CTGGCCTGTG	30	6	-	120	3.80	Homo sapiens mRNA for CIRP, complete cds
1352	CTGGCCTGTG	30	6	-	120	3.80	Villin 2 (ezrin)
1353	CTGGCCTGTG	30	6	-	120	3.80	Homo sapiens clone 23585 unknown mRNA, partial cds
1354	CACCCOCAGG	29	7	-	118	3.80	ESTs
1355	CACCCOCAGG	29	7	-	118	3.80	Human Gps2 (GPS2) mRNA, complete cds
1356	GTGAAACTCC	66	16	-	269	3.81	Human 53K isoform of Type II phosphatidylinositol-4-phosphate 5-kinase (PIPK) mRNA, complete cds
1357	GTGAAACTCC	66	16	-	269	3.81	Human mRNA for KIAA0328 gene, partial cds
1358	AGAATTGCTT	50	12	-	201	3.81	Homo sapiens nephrin (NPHS1) mRNA, complete cds
1359	AGAATTGCTT	50	12	-	201	3.81	H.sapiens mRNA for phosphorylase-kinase, beta subunit
1360	ATGGCCTCCT	19	5	-	76	3.84	Human syntaxin mRNA, complete cds
1361	AACTGTCCTT	34	5	-	138	3.84	H.sapiens mRNA for major astrocytic phosphoprotein PEA-15
1362	AAGGAATCGG	34	5	-	136	3.85	PROTEASOME BETA CHAIN PRECURSOR
1363	TCTGTTTATC	29	8	-	119	3.86	Signal recognition particle 14 kD protein
1364	ACTTTTTCAA	704	20	-	2741	3.87	Tag matches mitochondrial sequence
1365	TCTGTAATCC	46	8	-	185	3.87	Tag matches mitochondrial sequence
1366	TCTGTAATCC	46	8	-	185	3.87	Human aryl sulfotransferase mRNA, complete cds
1367	GTGAAACCC	27	5	-	110	3.90	No match
1368	GGCAGGCACA	24	5	-	97	3.91	H.sapiens mRNA for phenylalkylamine binding protein
1369	GGGGCAGGGC	281	33	-	1138	3.93	ESTs, Weakly similar to EPIDERMAL GROWTH FACTOR PRECURSOR, KIDNEY
1370	GGGGCAGGGC	281	33	-	1138	3.93	Eukaryotic translation initiation factor 5A
1371	GTGAAACTCT	32	8	-	134	3.94	No match
1372	TGGACGAGGC	26	7	-	118	3.95	ESTs, Weakly similar to No definition line found [C.elegans]
1373	CCTATAATCC	109	16	-	452	4.01	Retinoblastoma-like 1 (p107)
1374	CCTATAATCC	109	16	-	452	4.01	Cyclic nucleotide gated channel (photoreceptor), cGMP gated 2 (beta)
1375	CCTATAATCC	109	16	-	452	4.01	Homo sapiens mRNA for KIAA0694 protein, complete cds
1376	AACTGCTTCA	77	12	-	323	4.05	Homo sapiens Arp2/3 protein complex subunit p41-Arc (ARCA1) mRNA, complete cds
1377	GGATTGTCTG	55	11	-	233	4.07	Small nuclear ribonucleoprotein polypeptides B and B1
1378	CCTGTAATTC	48	8	-	201	4.07	Homo sapiens mRNA for KIAA0591 protein, partial cds
1379	CTGGGCCTGG	84	7	-	351	4.07	Human HU-K4 mRNA, complete cds
1380	ACCCTTGGCC	551	63	-	2334	4.08	Tag matches mitochondrial sequence
1381	ATGGCGATCT	27	7	-	117	4.09	Ribosomal protein S24
1382	TTGCTGCCT	39	8	-	166	4.10	ESTs
1383	TGAATCTGGG	35	6	-	150	4.11	SET translocation (myeloid leukemia-associated)
1384	AGCCTTTGTT	57	6	-	240	4.13	Human mRNA for collagen binding protein 2, complete cds
1385	CTTTTCAGCA	28	9	-	129	4.17	Human 14-3-3 epsilon mRNA, complete cds

WO 01/38577

PCT/US00/31922

Table 4, cont.

1386	CCTGGAGTGG	28	5	-	123	4.17	ESTs
1387	CGGAGACCT	87	14	-	380	4.20	Homo sapiens dbpB-like protein mRNA, complete cds
1388	CCTGGGTTT	1027	93	-	4414	4.21	Ferritin, light polypeptide
1389	ATTGACAAG	843	93	-	2814	4.23	Tag matches mitochondrial sequence
1390	ACAACCTCAAT	61	6	-	265	4.24	ESTs, Highly similar to BRAIN PROTEIN I3 [Mus musculus]
1391	CTTGATCCC	45	8	-	202	4.30	Homo sapiens quiescin (Qc) mRNA, complete cds
1392	GGCTGGTCTC	48	9	-	216	4.32	ESTs
1393	AGGTGGCAAG	194	45	-	891	4.36	Tag matches mitochondrial sequence
1394	CTAGCTTTTA	46	10	-	210	4.36	Tag matches mitochondrial sequence
1395	TCACCGGTCA	143	23	-	648	4.38	GELSOLIN PRECURSOR, PLASMA
1396	GGCCGGCTTC	110	5	-	487	4.38	Ribosomal protein S17
1397	GAGAGCTCCC	84	6	-	290	4.41	Tag matches mitochondrial sequence
1398	GAGAGCTCCC	84	6	-	290	4.41	EST
1399	GAGAGCTCCC	64	6	-	290	4.41	ESTs
1400	GAGAGCTCCC	64	6	-	290	4.41	Homo sapiens clone 24751 unknown mRNA
1401	CCCCGTACAT	122	7	-	849	4.43	No match
1402	TGGCGTACGG	87	11	-	314	4.50	Tag matches ribosomal RNA sequence
1403	TCCCGGACAT	97	5	-	444	4.53	No match
1404	CCTGGCTAAT	32	11	-	155	4.53	No match
1405	TCACAGGTGT	50	10	-	238	4.61	B-cell translocation gene 1, anti-proliferative
1406	TCCCATTAAG	119	12	-	560	4.61	No match
1407	TGCGACTGAG	259	21	-	1228	4.65	Major histocompatibility complex, class I, C
1408	GTGCACTGAG	259	21	-	1228	4.65	MHC class I protein HLA-A (HLA-A28, B40, -Cw3)
1409	GCTTACCTTT	35	6	-	170	4.68	Homo sapiens calumenin (Calu) mRNA, complete cds
1410	CTGGCCGGGA	84	7	-	264	4.71	Vasodilator-stimulated phosphoprotein
1411	CTGGCCGGGA	54	7	-	264	4.71	Homo sapiens Sox-like transcriptional factor mRNA, complete cds
1412	GGCCCTGTGC	133	11	-	647	4.79	Homo sapiens monocarboxylate transporter (MCT3) mRNA, complete cds
1413	GGCCCTGTGC	133	11	-	647	4.79	ESTs
1414	GCCCCTCCGG	121	18	-	598	4.79	ESTs, Weakly similar to TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICPO
1415	TTGTGATGTA	21	5	-	109	4.87	Neurotrophic tyrosine kinase, receptor, type 1
1416	TTGTGATGTA	21	5	-	109	4.87	Fibroblast growth factor receptor 4
1417	CATCTTCACC	62	5	-	311	4.97	Ribosomal protein S25
1418	TGGCCAGGA	100	35	-	539	5.06	No match
1419	AGAATCACTT	37	5	-	194	5.09	No match
1420	TAGCCAGGA	23	8	-	129	5.22	Human LLGL mRNA, complete cds
1421	GTGTGGTTA	486	43	-	2846	5.25	BETA-2-MICROGLOBULIN PRECURSOR
1422	CAAGCATCCC	547	36	-	2910	5.26	Tag matches mitochondrial sequence
1423	GACATATGTA	39	8	-	217	5.29	Cytochrome c oxidase subunit VIIb
1424	AGTATCTGGG	63	8	-	337	5.29	Homo sapiens Arp2/3 protein complex subunit p41-Arc (ARC41) mRNA, complete cds
1425	ACGCCCTGTG	120	19	-	659	5.35	Human transcriptional activator mRNA, complete cds
1426	CTCTTCGAGA	177	15	-	983	5.35	Glutathione peroxidase 1
1427	ATGAGCTGAC	104	11	-	571	5.42	CYSTATIN B
1428	GCCTCTGTCT	36	5	-	202	5.43	Ribosomal protein, large, P1
1429	AAGGAAGATC	38	6	-	214	5.43	Human glutathione-S-transferase homolog mRNA, complete cds
1430	AAACATCTCT	306	30	-	1698	5.45	Tag matches mitochondrial sequence
1431	CTCAGACAGT	84	5	-	385	5.95	ESTs, Highly similar to 40S RIBOSOMAL PROTEIN S27 [Rattus norvegicus]
1432	CCCAAGCTAG	435	54	-	2698	6.08	Heat shock 27kD protein 1

WO 01/38577

PCT/US00/31922

Table 4, cont.

1433	CCCAAGCTAG	435	54	2698	6.08	Tag matches ribosomal RNA sequence
1434	TCAATCAAGA	34	8	236	6.67	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide
1435	TGCAGCGCT	111	9	762	6.80	H.sapiens mRNA for uridine phosphorylase
1436	TTCACGTGTA	223	7	1557	6.94	Lectin, galactoside-binding, soluble, 3 (galectin 3) (NOTE: redefinition of symbol)
1437	CTGACCTGTG	228	16	1683	7.38	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 ALPHA CHAIN PRECURSOR
1438	GGGGTCAGGG	118	9	882	7.43	Glycogen phosphorylase B (brain form)
1439	GGCTTTAGGG	125	10	1019	8.05	Tag matches mitochondrial sequence
1440	TGGGTGAGCC	304	45	2536	8.21	Cathepsin B
1441	AGGGTGTTT	78	6	668	8.43	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase
1442	AGGGTGTTT	78	8	668	8.43	Tag matches mitochondrial sequence
1443	TGGTGTATGC	93	6	810	8.62	Tag matches mitochondrial sequence
1444	GAGTAGAGAA	50	8	465	9.15	SET translocation (myeloid leukemia-associated)
1445	TGCAGGCCTG	115	11	1185	10.02	TRYPTOPHANYL-TRNA SYNTHETASE
1446	GGGAAACCT	210	34	2242	10.51	V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3 (alternative products)
1447	GTGACCAACGG	4374	29	47260	10.80	Human N-methyl-D-aspartate receptor 2C subunit precursor (NMDAR2C) mRNA, complete cds
1448	GTGACCAACGG	4374	29	47260	10.80	Tag matches ribosomal RNA sequence

WO 01/38577

PCT/US00/31922

Table 5. Transcripts uniformly elevated in cancer tissues

Tag Sequence	SEQ ID NO:	Cancer tissues					Normal Tissues					Avg	UniGene Description
		CC	BC	BrC	LC	M	NC	NB	NBr	NL	NM		
ATGTGTAACG	226	93	72	13	5	48	0	0	3	0	0	30	S100 calcium-binding protein A4 (calcium protein, calvasculin, metastasin)
CCCTGCCCTG	227	53	86	120	56	20	21	27	0	6	0	21	Midline (neurile growth-promoting factor 2)
GTGCGCTGAG	228	85	103	360	23	56	0	30	56	0	6	18	Major histocompatibility complex, class I, C
CTGGCGCGTC	229	26	19	53	16	25	3	1	0	0	5	14	Apoptosis Inhibitor 4 (survivin)
GGCCCCCGCT	230	38	40	54	31	29	9	7	3	3	0	12	ESTs
TGCCCCACAG	231	13	201	8	24	336	0	30	3	3	19	9	Apolipoprotein C1
CCCTGTGTGG	232	16	14	17	16	6	0	0	0	0	3	9	ESTs
AGTGACCGAA	233	5	8	37	8	7	0	1	0	3	0	8	ESTs
CTGCACTTAC	234	52	34	81	64	76	3	12	22	5	30	6	DNA REPLICATION LICENSING FACTOR CDC47 HOMOLOG
CTGGCGAGCG	235	168	137	290	73	176	9	21	64	13	60	8	Human ubiquitin carrier protein (E2-EPPF) mRNA, complete cds
TTCGCGGTGC	236	4	10	12	19	7	0	1	0	0	0	7	ESTs
TGCGCTGGCC	237	22	63	74	28	14	6	18	6	8	0	7	No match
CTGCTGAAC	238	20	10	26	18	16	3	4	0	8	5	6	ESTs, Highly similar to MYO-INOSITOL-1-PHOSPHATE SYNTHASE [Arabidopsis thaliana]
CGCCCGTGT	239	4	151	30	9	30	0	13	6	0	5	6	No match
TTCGCCCGCT	240	10	61	15	19	23	0	22	6	5	0	6	AXL receptor tyrosine kinase
TTCCTAAGG	241	8	8	16	18	22	3	0	3	8	0	6	ESTs, Weakly similar to KIAA0005 [H.sapiens]
AGCCACGTTG	242	13	8	11	11	6	0	0	0	0	3	6	Add phosphatase 1, soluble
CTCTGGCACT	243	14	6	23	22	8	3	1	3	3	0	6	ESTs, Highly similar to transcription factor ARF6 chain B [M.musculus]
GCGCTCACT	244	23	13	52	16	17	3	4	6	3	5	6	Homo sapiens clone 24767 mRNA sequence / ESTs, Weakly similar to cdt [D.melanogaster]
CTTACAGCCA	245	11	6	19	12	6	0	0	3	0	3	6	ESTs
AGGCCCTCA	246	14	6	15	5	4	0	3	0	0	0	6	Homo sapiens mRNA, complete cds
GGGTAATGTG	247	7	13	5	11	12	0	1	0	0	5	5	ESTs, Moderately similar to unknown [M.musculus]
CTGACAGCCC	248	4	5	17	7	9	0	1	0	0	3	5	Human mRNA for Histone6, complete cds
TGACCTCCAG	249	7	14	15	12	11	0	8	3	3	0	5	ESTs, Weakly similar to No definition line found [C.elegans] / ESTs
AAAGCTCTTC	250	10	5	12	11	6	0	1	3	0	3	5	ESTs, Highly similar to G2M/TOTIC-SPECIFIC CYCLIN B2 [Mesocricetus auratus]
TCATTGCAT	251	7	13	5	4	9	3	1	0	0	0	5	ESTs, Highly similar to HYPOTHETICAL 16.3 KD PROTEIN [Saccharomyces cerevisiae]
CCCCCTCCGG	252	31	14	73	38	58	15	3	6	19	11	5	Small nuclear ribonucleoprotein polypeptide N / B and B1
GTAGGGGCT	253	11	14	11	19	18	3	6	0	3	8	4	ESTs
GAACCCAAAG	254	7	8	12	8	10	0	0	3	3	3	4	Plasminogen / PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A
TGTAGGCTC	255	5	11	11	7	7	0	3	0	0	3	4	Cyclin F
ATCTCTGGAG	256	7	3	9	8	7	0	0	0	0	3	4	ESTs
AAAGTGATC	257	10	19	11	4	7	0	9	0	0	3	4	No match
GCCTTGGGTG	258	7	8	4	9	10	3	3	0	0	0	4	Leukemia Inhibitory factor (cholinergic differentiation factor)
ACCTCACTCT	259	9	3	12	16	9	0	0	6	3	3	4	ESTs
TAAAGACTTG	260	9	13	24	12	38	3	1	11	5	11	4	Adenylate kinase 2 (adk2)
TGCGCGCCGG	261	15	16	21	14	6	6	3	6	3	0	4	SET translocation (myeloid leukemia-associated)
AACCTCAAGT	262	6	10	7	8	11	0	4	0	3	3	4	ESTs, Moderately similar to putative [M.musculus]
GTTTACCCGC	263	6	3	4	7	4	0	0	0	0	0	3	No match
GCCTCTGGCT	264	4	5	5	5	6	0	0	0	0	3	3	ESTs
CCTGGGTCT	265	4	10	6	5	7	0	4	3	0	3	3	ESTs

WO 01/38577

PCT/US00/31922

Table 6. Transcripts expressed in Colon Cancer Cells (>500 copies per cell)

Tag	SEQ ID NO:	Copies/cell	Unigene Description
CCCATCGTCC	1449	2672	Tag matches mitochondrial sequence
TGTGTGAGA	1450	1672	Translation elongation factor 1-alpha-1
GGATTGGCC	1451	1663	Ribosomal protein, large P2 / Ribosomal protein S26 / Human mRNA for PIG-B, complete cds
CCCGTCCGA	1452	1559	60S RIBOSOMAL PROTEIN L13
ATGGTGGTA	1453	1555	40S RIBOSOMAL PROTEIN S2
GTGAACCCC	1454	1482	Multiple matches
CCTCCAGTA	1455	1468	Keratin 8
TGGTCTCT	1456	1453	60S RIBOSOMAL PROTEIN L41
TGATTCACT	1457	1434	EST / Tag matches mitochondrial sequence
CCTGTAATCC	1458	1372	Multiple matches
ACTTTTCAA	1459	1367	Tag matches mitochondrial sequence
AAAAAAAAA	1460	1357	Multiple matches
GAGGGAGTT	1461	1290	Ribosomal protein L27a
GCGGAGGAG	1462	1141	Human mRNA for ribosomal protein S12
CACCTAATG	1463	1137	Tag matches mitochondrial sequence
CGCCGCCGC	1464	1098	Human ribosomal protein L35 mRNA, complete cds
GGGGAATCG	1465	1092	THYMOSIN BETA-10
GAAAAATGGT	1466	1056	Laminin receptor (2H5 epitope)
GGGCTGGGT	1467	1028	H. sapiens mRNA for ribosomal protein L29 / Homo sapiens sperm acrosomal protein mRNA
GCGGGTGGG	1468	986	Basigin
AGCCCTACAA	1469	945	Tag matches mitochondrial sequence
CTGGGTTAAT	1470	943	40S RIBOSOMAL PROTEIN S19
CAAACCATCC	1471	927	Keratin 18
TGCACGTTT	1472	916	Human mRNA for antileukoprotease (ALP) from cervix uterus
AGGCTACGGA	1473	905	60S RIBOSOMAL PROTEIN L13A
GCAGCCATCC	1474	861	Ribosomal protein L28
CTAATAAAA	1475	851	Ribosomal protein, large, P1 / TRANSCOBALAMIN1 PRECURSOR
CTAAGACTTC	1476	833	Tag matches mitochondrial sequence
TGGTGTGAG	1477	830	Human DNA sequence from clone 1033B10 on chromosome 6p21.2-21.31
TACCATCAAT	1478	828	Glyceraldehyde-3-phosphate dehydrogenase
TTCATACACC	1479	814	Tag matches mitochondrial sequence
CGACTGCACT	1480	800	Multiple matches
ACTAACACCC	1481	795	Tag matches mitochondrial sequence
AAGGTGGAGG	1482	794	60S RIBOSOMAL PROTEIN L18A
AGCACCTCCA	1483	787	Eukaryotic translation elongation factor 2
CACAAACGGT	1484	761	40S RIBOSOMAL PROTEIN S27
AGGAAAGCTG	1485	732	EST's, highly similar to 60S RIBOSOMAL PROTEIN L36 [Rattus norvegicus]
GTGAACCCCT	1486	729	Multiple matches
AATCCGTGG	1487	711	Ribosomal protein L8
TGGGGTTTC	1488	698	Ferritin heavy chain
AAGACAGTGG	1489	656	Ribosomal protein L37a
ATTGAGAAG	1490	640	Tag matches mitochondrial sequence
GCGGTGTCGG	1491	679	Human ribosomal protein S6 mRNA, complete cds

WO 01/38577

PCT/US00/31922

Table 6, cont.

CGCCGGAACA	1492	Ribosomal protein L4	678
TCTCATACC	1493	Tag matches mitochondrial sequence	661
ACATCATCGA	1494	Ribosomal protein L12	661
AAGCGGCCA	1495	Macrophage migration inhibitory factor	644
AGGCTTCCA	1496	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX SUBUNIT VI REQUIRING PROTEIN	643
CCGTCCAAGG	1497	Ribosomal protein S16	631
CGCTGGTCC	1498	Homo sapiens ribosomal protein L11 mRNA, complete cds	626
CTCAACATCT	1499	Ribosomal protein, large, P0	615
ACTCCAAAA	1500	H sapiens mRNA for transmembrane protein mp24 / Human insulinoma rig-analog mRNA encoding DNA-binding protein	608
CCTAGCTGGA	1501	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A	606
GTGAAGGCAG	1502	Ribosomal protein S3A	596
AGCTCTCCCT	1503	60S RIBOSOMAL PROTEIN L23	551
TAGGTTGTCT	1504	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN	537
GGACCACTGA	1505	Ribosomal protein L3	522
AAGGAGATGG	1506	Ribosomal protein L31	521
AACTAAAAA	1507	Ubiquitin A-52 residue ribosomal protein fusion product 1	510
GGCTGGGGGC	1508	Human profilin mRNA, complete cds	507
CCAGAACAGA	1509	Deoxythymidylate kinase / 60S RIBOSOMAL PROTEIN L30	503

WO 01/38577

PCT/US00/31922

Table 7. Expressed transcripts (>500 copies per cell)

Tag Sequence	SEQ ID NO:	Copies/Cell	Description
CCCATCGTCC	1508	3022	Tag matches mitochondrial sequence
GTGACCAACGG	1509	2435	Tag matches ribosomal RNA sequence / Human N-methyl-D-aspartate receptor 2C subunit precursor (NMDAR2C) mRNA
TGTGTTGAGA	1510	1557	Translation elongation factor 1-alpha-1
GTGAACCC	1511	1456	Multiple matches
CCTGTAATCC	1512	1403	Multiple matches
CTAAGACTTC	1513	1349	Tag matches mitochondrial sequence
CACGTAATTG	1514	1333	Tag matches mitochondrial sequence
CCCGTCGGGA	1515	1282	60S RIBOSOMAL PROTEIN L13
TTGGTCCCTCT	1516	1238	60S RIBOSOMAL PROTEIN L41
ATGGCTGGTA	1517	1126	40S RIBOSOMAL PROTEIN S2
TTGGGGTTTC	1518	1099	Ferritin heavy chain
CCACTGCACT	1519	964	Multiple matches
TGATTCACT	1520	942	Tag matches mitochondrial sequence / EST
ACTTTTCAA	1521	899	Tag matches mitochondrial sequence
GCAGCCATCC	1522	886	Ribosomal protein L28
TACCATCAAT	1523	874	Glyceraldehyde-3-phosphate dehydrogenase
GGATTGGCC	1524	854	Ribosomal protein, large P2 / Ribosomal protein S26 / Human mRNA for PIG-B
CCCTGGGTTTC	1525	844	Ferritin, light polypeptide
GCCGAGGAAG	1526	836	Human mRNA for ribosomal protein S12
AGGCTACGGA	1527	820	60S RIBOSOMAL PROTEIN L13A
CGCCGCCGGC	1528	805	Human ribosomal protein L35 mRNA, complete cds
TTCATAGACC	1529	804	Tag matches mitochondrial sequence
AGCCCTACAA	1530	801	Tag matches mitochondrial sequence
CACAAACGGT	1531	799	40S RIBOSOMAL PROTEIN S27
AAGGTGGAGG	1532	786	60S RIBOSOMAL PROTEIN L18A
CTTCTTGCC	1533	777	Keratin 17
TGGTGTGAG	1534	770	Human DNA sequence from clone 1033B10 on chromosome 6p21.2-21.31
GTGAACCCCT	1535	728	Multiple matches
GGGGAATCG	1536	724	THYMOSIN BETA-10
AGCACCTCCA	1537	718	Eukaryotic translation elongation factor 2
CCTCCAGCTA	1538	711	Keratin 8
AAGACAGTGG	1539	699	Ribosomal protein L37a
CTGGGTAAAT	1540	699	40S RIBOSOMAL PROTEIN S19
ATTTGAGAAG	1541	689	Tag matches mitochondrial sequence
GCCGGGTGGG	1542	687	Basigin
GGGCTGGGGT	1543	683	H.sapiens mRNA for ribosomal protein L29 / Homo sapiens sperm acrosomal protein mRNA
AGGGCTTCCA	1544	663	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX SUBUNIT VI REQUIRING PROTEIN
AAAAAATAA	1545	650	Multiple matches
GAGGGAGTTT	1546	648	Ribosomal protein L27a
GCGACCGTCA	1547	637	Aldolase A
AGTAACACCC	1548	631	Tag matches mitochondrial sequence
CGCCGGGAAC	1549	616	Ribosomal protein L4
TGGGCAAGC	1550	592	Translation elongation factor 1 gamma
TGCACGTTTT	1551	586	Human mRNA for antileukoprotease (ALP) from cervix uterus

WO 01/38577

PCT/US00/31922

Table 7, cont.

AATCCTGTGG	1552	569	Ribosomal protein L8
CAAGCATCCC	1553	565	Tag matches mitochondrial sequence
CCGTCCAAGG	1554	559	Ribosomal protein S16
TAGGTGTCT	1555	551	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
GCCGTGTCCG	1556	540	Human ribosomal protein S6 mRNA, complete cds
GCTTTATTG	1557	540	Human mRNA fragment encoding cytoplasmic actin
CTAGCCTCAC	1558	539	Actin, gamma 1
GCTAGCTGGA	1559	537	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A
GCCCTGTGCTG	1560	534	Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne types)
ACCCTTGGCC	1561	526	Tag matches mitochondrial sequence
AGGAAGCTG	1562	513	ESTs, Highly similar to 60S RIBOSOMAL PROTEIN L36 [Rattus norvegicus]

WO 01/38577

PCT/US00/31922

CLAIMS

1. A method of identifying a cell as either a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, or a kidney epithelial cell, comprising the step of:

determining expression in a test cell of a gene product of at least one gene comprising a sequence selected from at least one of the following groups:

- (a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;
- (b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, and 131-150, and 151;
- (c) the sequences shown in SEQ ID NOS:152-154, and 155;
- (d) the sequences shown in SEQ ID NOS:156-159, and 160;
- (e) the sequences shown in SEQ ID NOS:161-166, and 167;
- (f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;
- (g) the sequences shown in SEQ ID NOS:209 and 210; and
- (h) the sequences shown in SEQ ID NOS:211-224 and 225,

wherein expression of a gene product of at least one gene comprising a sequence shown in (a) identifies the test cell as a colon epithelial cell;

wherein expression of a gene product of at least one gene comprising a sequence shown in (b) identifies the test cell as a brain cell;

wherein expression of a gene product of at least one gene comprising a sequence shown in (c) identifies the test cell as a keratinocyte;

wherein expression of a gene product of at least one gene comprising a sequence shown in (d) identifies the test cell as a breast epithelial cell;

wherein expression of a gene product of at least one gene comprising a sequence shown in (e) identifies the test cell as a lung epithelial cell;

wherein expression of a gene product of at least one gene comprising a sequence shown in (f) identifies the test cell as a melanocyte;

wherein expression of a gene product of at least one gene comprising a sequence shown in (g) identifies the test cell as a prostate cell; and

WO 01/38577

PCT/US00/31922

wherein expression of a gene product of at least one gene comprising a sequence shown in (h) identifies the test cell as a kidney epithelial cell.

2. The method of claim 1 wherein expression of gene products of at least two of said genes is determined.

3. The method of claim 1 wherein expression of gene products of at least five of said genes is determined.

4. The method of claim 1 wherein the gene product is protein.

5. The method of claim 1 wherein the gene product is RNA.

6. The method of claim 5 wherein expression is determined using at least one oligonucleotide probe.

7. The method of claim 5 wherein expression is determined using at least two oligonucleotide probes.

8. The method of claim 6 wherein the at least one oligonucleotide probe is immobilized on a solid support.

9. The method of claim 8 wherein the at least one oligonucleotide probe is in an array.

10. The method of claim 1 wherein the cell to be identified is a cancer cell.

11. An isolated polynucleotide comprising a sequence selected from the group consisting of SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-84, 98, 103, 113, 115, 122, 129, 132, 134, 135, 140, 144, 149, 150, 153-168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-213, 216-224, 237, 239, 257, 263, 485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

12. A solid support comprising at least one polynucleotide comprising a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5, 6, 8, 10, 12, 13, 15, 17, 18, 21, 24-26, 28, 30, 31, 34-36, 38, 40, 47-51, 53-57, 59-62, 65-69, 71-76, 78, 80-83, and 84;

(b) the sequences shown in SEQ ID NOS:98, 103, 113, 115, 122, 129, 132,

WO 01/38577

PCT/US00/31922

134, 135, 140, 144, 149, and 150;

(c) the sequences shown in SEQ ID NOS:153-154 and 155;

(d) the sequences shown in SEQ ID NOS:156-157 and 160;

(e) the sequences shown in SEQ ID NOS:161-166 and 167;

(f) the sequences shown in SEQ ID NOS:168, 174-176, 182, 185, 186, 188, 190, 200, 201, 205-207 and 208;

(g) the sequences shown in SEQ ID NOS:209 and 210;

(h) the sequences shown in SEQ ID NOS:211-213, 216-223, and 224;

(i) the sequences shown in SEQ ID NOS:237, 239, 257, and 263; or

(j) the sequences shown in SEQ ID NOS:485, 487, 495, 499, 514, 586, 686, 751, 835, 844, 878, 910, 925, 932, 951, 1000, 1005, 1070, 1122, 1130, 1170, 1173, 1187, 1189, 1200, 1213, 1220, 1237, 1257, 1264, 1273, 1293, 1300, 1320, 1367, 1371, 1401, 1403, 1404, 1406, 1418, and 1419.

13. The solid support of claim 12 wherein:

if the at least one polynucleotide comprises a sequence selected from (a), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:1, 3, 4, 7, 9, 11, 14, 16, 19, 20, 22, 23, 27, 29, 32, 33, 37, 39, 41-46, 52, 58, 63, 64, 70, 77, 79, and 85;

if the at least one polynucleotide comprises a sequence selected from (b), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:86-97, 99-102, 104-112, 114, 116-121, 123-128, 130, 131, 133, 136-139, 141-143, 145-148, and 151;

if the at least one polynucleotide comprises a sequence selected from (c), then the solid support further comprises a polynucleotide comprising the sequence shown in SEQ ID NO:152;

if the at least one polynucleotide comprises a sequence selected from (f), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:169-173, 177-181, 183, 184, 187, 189, 191-199, 202, 203, and 204;

if the at least one polynucleotide comprises a sequence selected from (h), then

WO 01/38577

PCT/US00/31922

the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:214, 215, and 225;

if the at least one polynucleotide comprises a sequence selected from (i), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:226-236, 238, 240-256, 258-262, 264, and 265; and

if the at least one polynucleotide comprises a sequence selected from (j), then the solid support further comprises a polynucleotide comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:266-484, 486, 488-494, 496-498, 500-513, 515-585, 587-685, 687-750, 752-834, 836-843, 845-877, 879-909, 911-924, 926-931, 933-950, 952-999, 1001-1004, 1006-1069, 1071-1121, 1123-1129, 1131-1169, 1171, 1171, 1174-1186, 1188, 1190-1199, 1201-1212, 1214-1219, 1221-1236, 1238-1256, 1258-1263, 1265-1272, 1274-1292, 1294-1299, 1301-1319, 1321-1366, 1368-1370, 1372-1400, 1402, 1405, 1407-1416, and 1417.

14. The solid support of claim 12, wherein:

if the at least one polynucleotide comprises a sequence selected from (a), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:1, 3, 4, 7, 9, 11, 14, 16, 19, 20, 22, 23, 27, 29, 32, 33, 37, 39, 41-46, 52, 58, 63, 64, 70, 77, 79, and 85;

if the at least one polynucleotide comprises a sequence selected from (b), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:86-97, 99-102, 104-112, 114, 116-121, 123-128, 130, 131, 133, 136-139, 141-143, 145-148, and 151;

if the at least one polynucleotide comprising a sequence selected from (c), then the at least one polynucleotide further comprises SEQ ID NO:152;

if the at least one polynucleotide comprises a sequence selected from (f), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:169-173, 177-181, 183, 184, 187, 189, 191-199, 202, 203, and 204;

if the at least one polynucleotide comprises a sequence selected from (h), then

WO 01/38577

PCT/US00/31922

the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:214, 215, and 225;

if the at least one polynucleotide comprises a sequence selected from (i), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:226-236, 238, 240-256, 258-262, 264, and 265; and

if the at least one polynucleotide comprises a sequence selected from (j), then the at least one polynucleotide further comprises a sequence selected from the group consisting of the sequences shown in SEQ ID NOS:266-484, 486, 488-494, 496-498, 500-513, 515-585, 587-685, 687-750, 752-834, 836-843, 845-877, 879-909, 911-924, 926-931, 933-950, 952-999, 1001-1004, 1006-1069, 1071-1121, 1123-1129, 1131-1169, 1171, 1171, 1174-1186, 1188, 1190-1199, 1201-1212, 1214-1219, 1221-1236, 1238-1256, 1258-1263, 1265-1272, 1274-1292, 1294-1299, 1301-1319, 1321-1366, 1368-1370, 1372-1400, 1402, 1405, 1407-1416, and 1417.

15. The solid support of claim 12 wherein the at least one polynucleotide is in an array.

16. A method of identifying a test cell as a cancer cell, comprising the step of:
determining expression in a test cell of a gene product of at least one gene comprising a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259-260, and 262-265, wherein an increase in said expression of at least two-fold relative to expression of the at least one gene in a normal cell identifies the test cell as a cancer cell.

17. The method of claim 16 wherein expression of gene products of at least two of said genes is determined.

18. The method of claim 16 wherein expression of gene products of at least five of said genes is determined.

19. The method of claim 16 wherein the gene product is protein.

20. The method of claim 16 wherein the gene product is RNA.

21. The method of claim 20 wherein expression is determined using at least one oligonucleotide probe.

22. The method of claim 21 wherein expression is determined using at least two

WO 01/38577

PCT/US00/31922

oligonucleotide probes.

23. The method of claim 21 wherein the at least one oligonucleotide probe is immobilized on a solid support.

24. The method of claim 23 wherein the at least one oligonucleotide probe is in an array.

25. The method of claim 16 wherein the test cell is selected from the group consisting of a colon epithelial cell, a breast epithelial cell, a lung epithelial cell, a melanocyte, and a brain cell.

26. The method of claim 16 wherein the normal cell and the test cell are selected from a single cell type.

27. A method of reducing expression of a cancer-specific gene in a human cell, comprising the step of:

administering to the cell a reagent which specifically binds to an expression product of a cancer-specific gene comprising a sequence selected from the group consisting of SEQ ID NOS:228, 230-257, 259-260, and 262-265, whereby expression of the cancer-specific gene is reduced relative to expression of the cancer-specific gene in the absence of the reagent.

28. The method of claim 27 wherein the reagent is an antisense oligonucleotide.

29. The method of claim 27 wherein the reagent is an antibody.

30. A method for comparing expression of a gene in a test sample to expression of a gene in a standard sample, comprising the steps of:

determining a first ratio and a second ratio, wherein the first ratio is an amount of an expression product of a test gene in a test sample to an amount of an expression product of at least one gene comprising a sequence selected from the group consisting of SEQ ID NOS:266-375, 377-652, 654-796, and 798-1448 in the test sample, and wherein the second ratio is an amount of an expression product of the test gene in a standard sample to an amount of an expression product of the at least one gene in the standard sample; and

comparing the first and second ratios, wherein a difference between the first and second ratios indicates a difference in the amount of the expression product of the test

WO 01/38577

PCT/US00/31922

gene in the test sample.

31. The method of claim 30 wherein the at least one gene comprises a sequence selected from the group consisting of SEQ ID NOS:282, 288, 300, 302, 308, 320, 323, 363, 368, 379, 381, 444, 453, 518, 531, 535, 538, 542, 579, 580, 594, 600, 604, 617, 626, 641, 650, 717, 728, 776, 777, 794, 818, 822, 842, 885, 887, 899, 900, 902, 904, 914, 930, 960, 964, 1001, 1015, 1020, 1027, 1035, 1090, 1113, 1119, 1146, 1151, 1163, 1233, 1235, 1252, 1255, 1270, 1340, 1345, 1356, 1359, 1360, 1362, 1385, 1415, and 1441.

32. The method of claim 30 wherein expression is determined using at least one oligonucleotide probe.

33. The method of claim 32 wherein the at least one oligonucleotide probe is immobilized on a solid support.

34. The method of claim 33 wherein the at least one oligonucleotide probe is in an array.

35. The method of claim 30 wherein the test sample is a cancer cell and the standard sample is a normal cell.

36. The method of claim 35 wherein the cancer cell is selected from the group consisting of a colon cancer cell, a breast cancer cell, a lung cancer cell, a melanoma cell, and a brain cancer cell.

37. The method of claim 30 wherein the test sample has been treated with a test compound and the standard sample has not been treated with the test compound.

38. The method of claim 37 wherein the test sample is a cancer cell and wherein the standard sample is a normal cell.

39. The method of claim 30 wherein the test sample and the standard sample are obtained from the same cell type.

40. A method of screening candidate anti-cancer drugs, comprising the steps of:
contacting a cancer cell with a test compound; and
measuring expression in the cancer cell of a gene product of at least one gene comprising a sequence selected from the group consisting of SEQ ID NOS: 228, 230-257, 259, 260, 262-263, and 265, wherein a decrease in expression of the gene product in the presence of a test compound relative to expression of the gene product in the absence of the

WO 01/38577

PCT/US00/31922

test compound identifies the test compound as a potential anti-cancer drug.

41. The method of claim 40 wherein the cancer cell is selected from the group consisting of a colon cancer cell, a breast cancer cell, a lung cancer cell, a melanoma cell, and a brain cancer cell.

42. The method of claim 40 in which expression of gene products of at least two of said genes is measured.

43. The method of claim 40 in which expression of gene products of at least five of said genes is measured.

44. The method of claim 40 wherein the gene product is protein.

45. The method of claim 40 wherein the gene product is RNA.

46. The method of claim 45 wherein expression of the at least one gene product is measured using at least one oligonucleotide probe.

47. The method of claim 46 wherein the at least one oligonucleotide probe is immobilized on a solid support.

48. The method of claim 47 wherein the at least one oligonucleotide probe is in an array.

49. The method of claim 46 wherein the at least one oligonucleotide probe comprises a sequence selected from the group consisting of SEQ ID NOS:237, 239, 257, and 263.

50. A method of screening test compounds for the ability to increase an organ or cell function, comprising the step of:

contacting a cell selected from the group consisting of a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, and a kidney cell with a test compound; and

measuring expression in the cell of a gene product of at least one gene comprising a sequence selected from at least one of the following groups:

(a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;

(b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;

(c) the sequences shown in SEQ ID NOS:152-154, and 155;

WO 01/38577

PCT/US00/31922

- (d) the sequences shown in SEQ ID NOS:156-159 and 160;
- (e) the sequences shown in SEQ ID NOS:161-166 and 167;
- (f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;
- (g) the sequences shown in SEQ ID NOS:209 and 210; and
- (h) the sequences shown in SEQ ID NOS:211-224 and 225,

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (a) identifies the test compound as a potential drug for increasing a function of a colon cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (b) identifies the test compound as a potential drug for increasing a function of a brain cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (c) identifies the test compound as a potential drug for increasing a function of a skin cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (d) identifies the test compound as a potential drug for increasing a function of a breast cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (e) identifies the test compound as a potential drug for increasing a function of a lung cell;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (f) identifies the test compound as a potential drug for increasing a function of a melanocyte;

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (g) identifies the test compound as a potential drug for increasing a function of a prostate cell; and

wherein an increase in expression of a gene product of at least one gene comprising a sequence selected from (h) identifies the test compound as a potential drug for increasing a function of a kidney cell.

WO 01/38577

PCT/US00/31922

51. The method of claim 50 wherein expression of gene products of at least two of said genes is determined.

52. The method of claim 50 wherein expression of gene products of at least five of said genes is determined.

53. The method of claim 50 wherein the gene product is protein.

54. The method of claim 50 wherein the gene product is RNA.

55. The method of claim 54 wherein expression is determined using at least one oligonucleotide probe.

56. The method of claim 54 wherein expression is determined using at least two oligonucleotide probes.

57. The method of claim 55 wherein the at least one oligonucleotide probe is immobilized on a solid support.

58. The method of claim 57 wherein the at least one oligonucleotide probe is in an array.

59. A method to restore function to a diseased tissue or cell comprising the step of:

delivering a gene to a diseased cell selected from the group consisting of a colon epithelial cell, a brain cell, a keratinocyte, a breast epithelial cell, a lung epithelial cell, a melanocyte, a prostate cell, and a kidney cell, wherein the gene comprises a nucleotide sequence selected from at least one of the following groups :

- (a) the sequences shown in SEQ ID NOS:2, 5-18, 20-84, and 85;
- (b) the sequences shown in SEQ ID NOS:87-96, 98, 100-103, 105, 107-110, 112-129, 131-150, and 151;
- (c) the sequences shown in SEQ ID NOS:152-154, and 155;
- (d) the sequences shown in SEQ ID NOS:156-159 and 160;
- (e) the sequences shown in SEQ ID NOS:161-166 and 167;
- (f) the sequences shown in SEQ ID NOS:168, 170, 172-177, 179-188, 190-207, and 208;
- (g) the sequences shown in SEQ ID NOS:209 and 210; and
- (h) the sequences shown in SEQ ID NOS:211-224 and 225,

WO 01/38577

PCT/US00/31922

wherein expression of the gene in the diseased cell is less than expression of the gene in a corresponding cell which is normal,

wherein if the diseased cell is a colon epithelial cell, then the nucleotide sequence is selected from (a);

wherein if the diseased cell is a brain cell, then the nucleotide sequence is selected from (b);

wherein if the diseased cell is a keratinocyte, then the nucleotide sequence is selected from (c);

wherein if the diseased cell is a breast epithelial cell, then the nucleotide sequence is selected from (d);

wherein if the diseased cell is a lung epithelial cell, then the nucleotide sequence is selected from (e);

wherein if the diseased cell is a melanocyte, then the nucleotide sequence is selected from (f);

wherein if the diseased cell is a prostate cell, then the nucleotide sequence is selected from (g); and

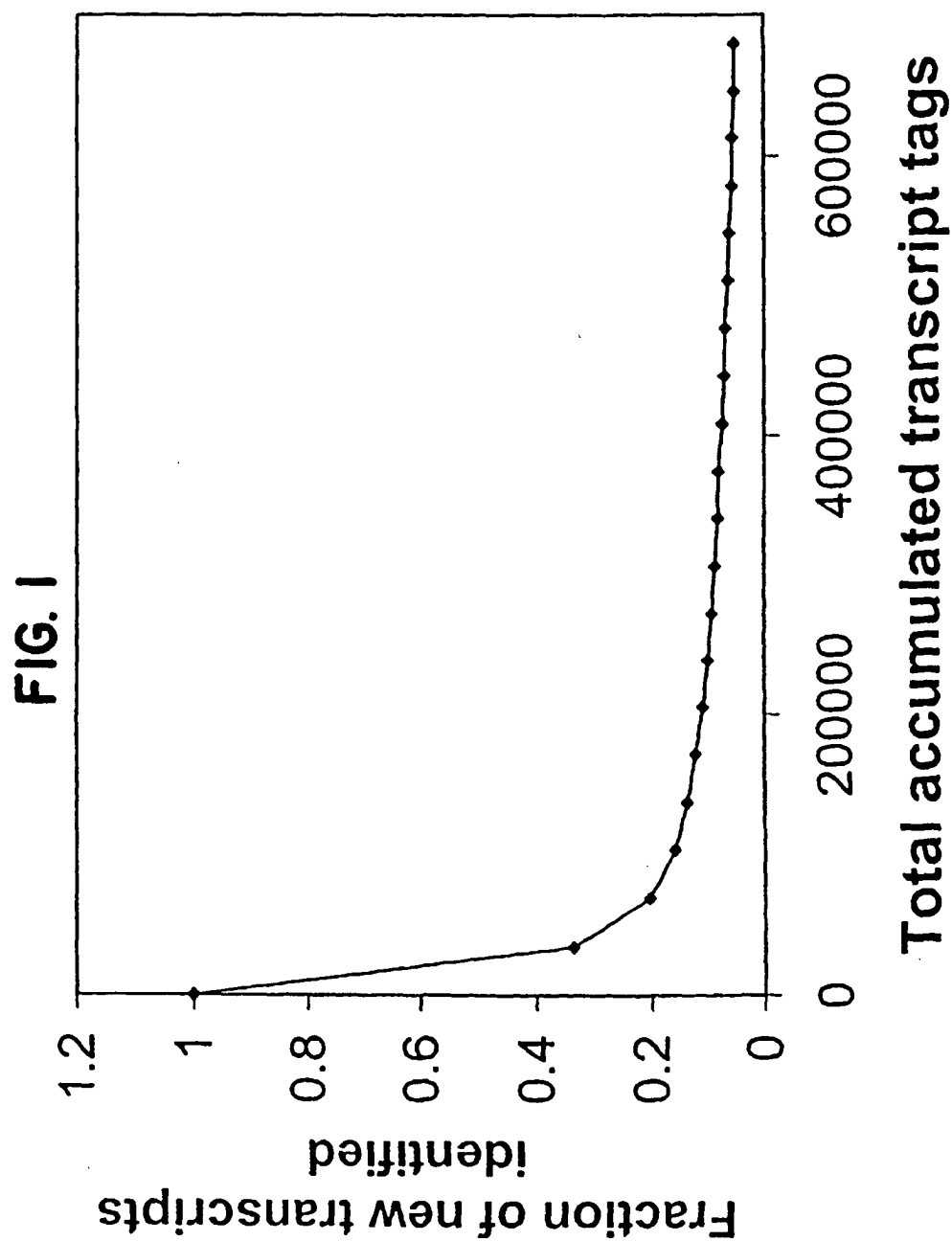
wherein if the diseased cell is a kidney cell, then the nucleotide sequence is selected from (h).

60. The method of claim 59 wherein the diseased cell fails to express the gene in the diseased state.

WO 01/38577

PCT/US00/31922

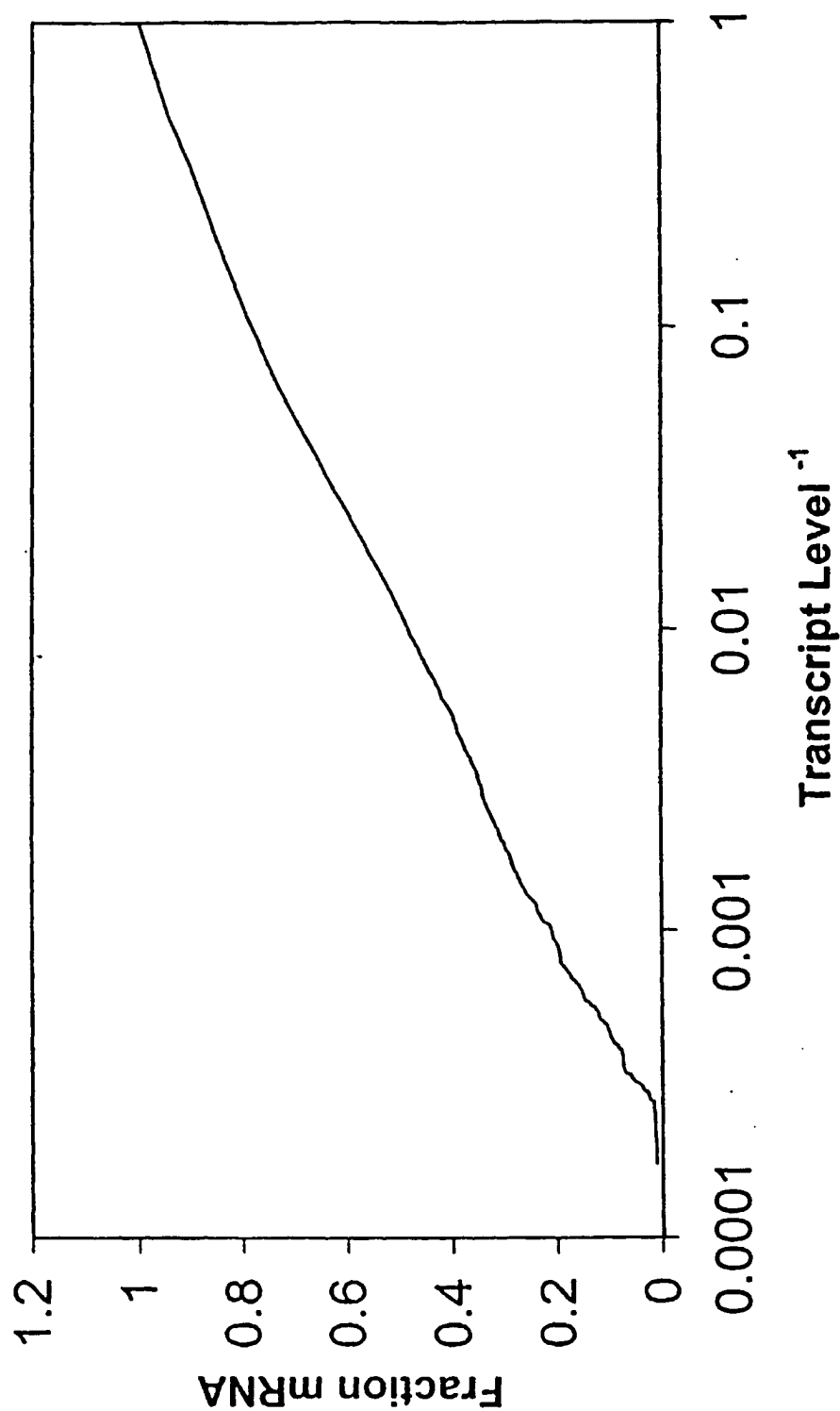
1/5



WO 01/38577

PCT/US00/31922

2/5

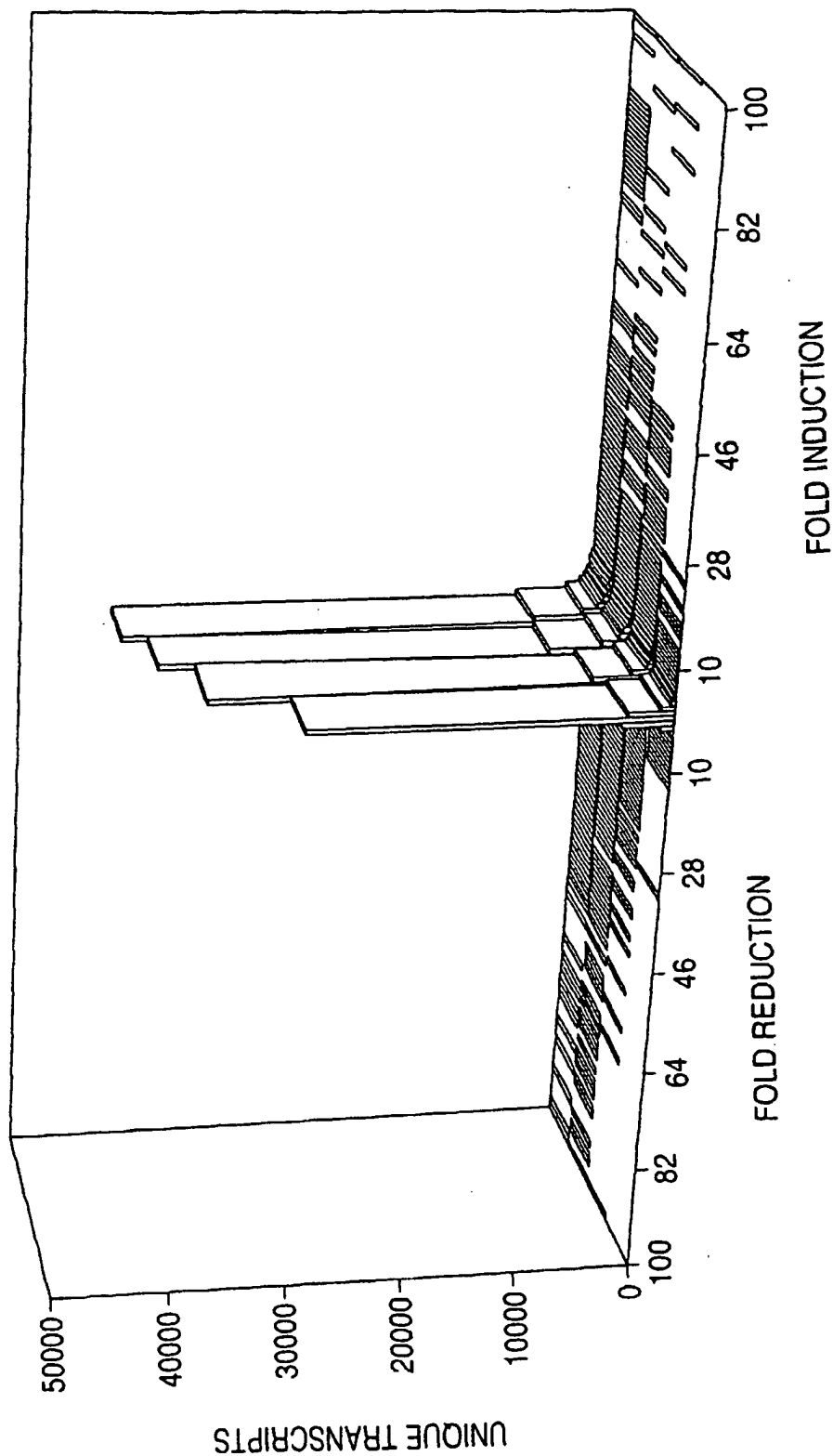
FIG. 2 Colon Cancer Cell Rot Curve

WO 01/38577

PCT/US00/31922

3/5

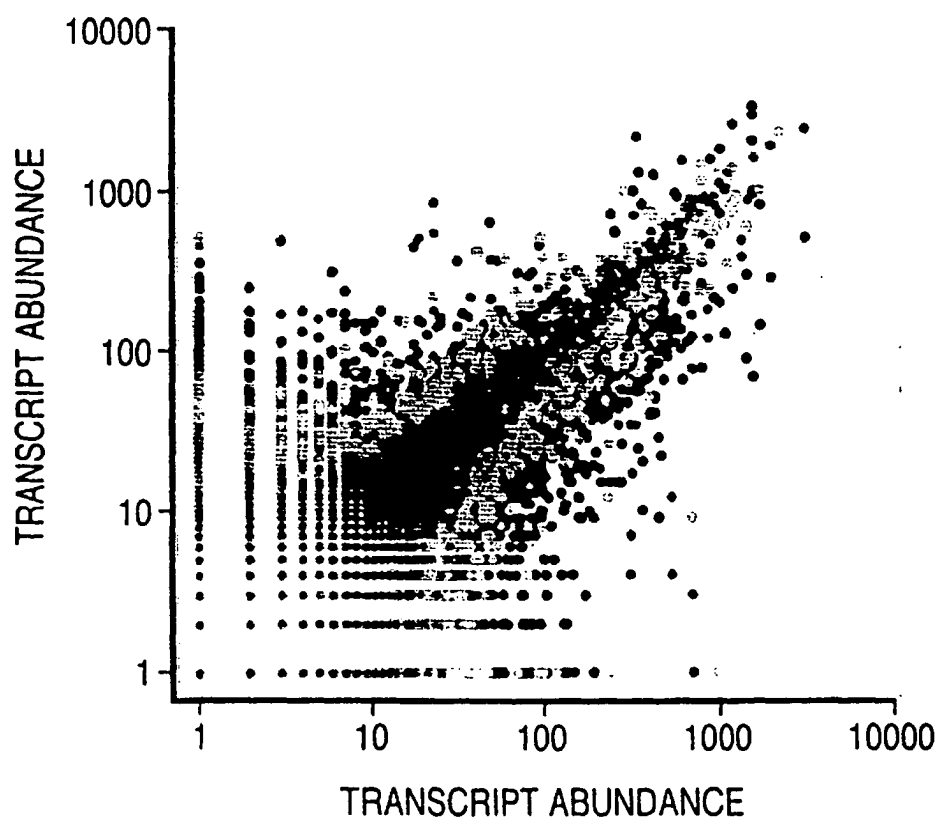
FIG. 3A



WO 01/38577

PCT/US00/31922

4/5

FIG. 3B

WO 01/38577

PCT/US00/31922

FIG. 3C

COMPARISON	DESCRIPTION	TOTAL TRANSCRIPTS	UNIQUE TRANSCRIPTS	EXPRESSION CHANGE > = 10 FOLD (%)
<input type="checkbox"/> 1	DLD1 CONDITION A vs DLD1 CONDITION B	0	42,673	43 (0.10)
<input type="checkbox"/> 2	DLD1 vs HCT116	0	56,061	390 (0.70) ^A
<input checked="" type="checkbox"/> 3	COLON CANCER vs NORMAL BRAIN	0	62,216	930 (1.49) ^B
<input checked="" type="checkbox"/> 4	COLON CANCER vs HEMANGIOPERICYTOMA	0	72,239	1,047 (1.45) ^C

5/5

^A DIFFERENCE BETWEEN EXPRESSION CHANGE OF COMPARISON 1 AND 2, $p < 0.0001$ ^B DIFFERENCE BETWEEN EXPRESSION CHANGE OF COMPARISON 2 AND 3, $p < 0.0001$ ^C DIFFERENCE BETWEEN EXPRESSION CHANGE OF COMPARISON 2 AND 4, $p < 0.0001$